

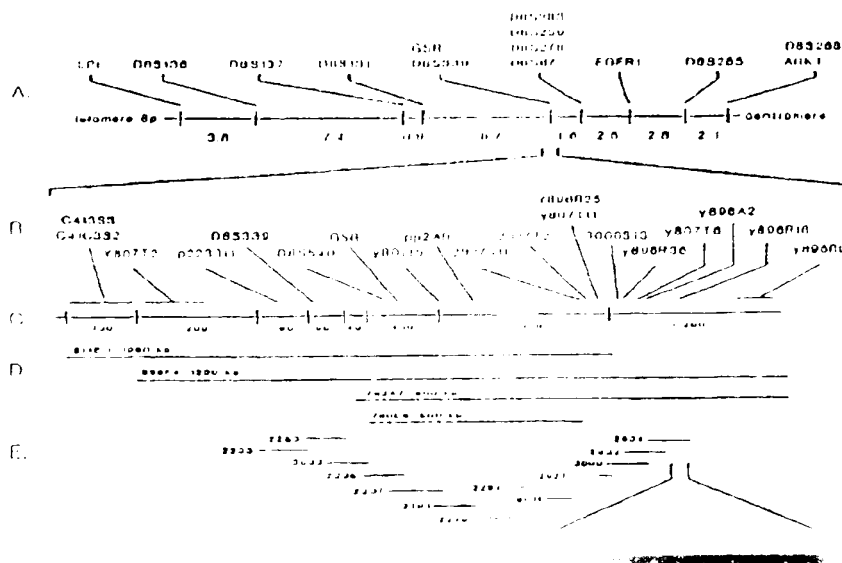
PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/00, 15/61, 9/90, 15/85, 15/86, 5/10, 5/20, C07K 16/40, C12Q 1/68, A01K 67/027, G01N 33/50		A1	(11) International Publication Number: WO 97/24435
(21) International Application Number: PCT/US96/20785		(43) International Publication Date: 10 July 1997 (10.07.97)	
(22) International Filing Date: 30 December 1996 (30.12.96)		(74) Agents: NOTTENBURG, Carol, Ph., D. et al.; Seed and Berry L.L.P., 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).	
(30) Priority Data: 60/009,409 29 December 1995 (29.12.95) US 08/580,539 29 December 1995 (29.12.95) US 60/010,835 30 January 1996 (30.01.96) US 08/594,242 30 January 1996 (30.01.96) US 08/632,175 12 April 1996 (12.04.96) US		(81) Designated States: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(71) Applicant: DARWIN MOLECULAR CORPORATION [US/US]; 1631-220th Street Southeast, Bothell, WA 98071 (US).		Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(71)(72) Applicant and Inventor: OSHIMA, Junko [US/US]; 5200 Ravenna Avenue N.E., Seattle, WA 98105 (US).			
(72) Inventors: FU, Ying-Hui; 7417-28th Avenue N.W., Seattle, WA 98117 (US); YU, Chang-En; 5709-60th Avenue N.E., Seattle, WA 98105 (US); MULLIGAN, John; 5823-17th Avenue Northeast, Seattle, WA 98105 (US); SCHELLEN- BERG, Gerald, D.; 7031-19th Avenue Northwest, Seattle, WA 98117 (US).			

(54) Title: GENES AND GENE PRODUCTS RELATED TO WERNER'S SYNDROME



Abstract

The present invention discloses nucleic acid molecules encoding WRN gene products, expression vectors and host cells suitable for expressing such products.

AL

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland			SI	Slovenia
CI	Ivory Coast				
CM	Cameroun				
CO	Colombia				
CR	Costa Rica				
CU	Cuba				
CY	Cyprus				
CZ	Czech Republic				
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MS	Montserrat	UZ	Uzbekistan
GA	Gabon	MR	Mauritania		

DESCRIPTION

GENES AND GENE PRODUCTS RELATED TO WERNER'S SYNDROME

5 TECHNICAL FIELD

The present invention relates generally to Werner's Syndrome and more specifically to methods and compositions suitable for use in diagnosis and treatment of Werner's Syndrome.

10 BACKGROUND OF THE INVENTION

Werner Syndrome (WS) is an autosomal recessive disorder with a complex phenotype. The disorder manifests itself in premature occurrence of age-related diseases and premature appearance of some of the physical features of normal aging. The onset of symptoms usually occurs after adolescence. The disorder progresses throughout life and typically patients have a shortened life expectancy with a modal age of death at 47. The prevalence of Werner Syndrome is estimated for heterozygotes to be 1-5 per 1,000 individuals, and for homozygotes to be 1-22 per 1,000,000 individuals.

Clinical symptoms of Werner Syndrome include both a prevalence of age-related diseases and physical features of aging. Such diseases include arteriosclerosis and heart disease, both benign and malignant neoplasms (usually sarcomas), diabetes mellitus, osteoporosis, and ocular cataracts. The physical appearance of WS patients is often manifest as a short stature, premature graying or loss of hair, hypogonadism, altered skin pigmentation, hyperkeratosis, tight skin, bird-like facies, cutaneous atrophy, cutaneous leg ulcers, and telangiectasia. Most of these diseases and features are present in from 40-90% of WS patients. Diagnosis of WS relies mainly upon the appearance of a certain number of these diseases and features

In addition to the noted signs and symptoms of aging, Werner Syndrome mimics normal aging as evidenced by the replicative potential of fibroblasts isolated from WS subjects. Replication potential of fibroblasts is reduced in these patients compared to fibroblasts isolated from age-matched controls, and is comparable to the replicative potential of fibroblasts taken from elderly subjects. Moreover, an increased mutation rate has been described in WS patients. Such abnormality is manifest as chromosomal instability, such as inversions, reciprocal translocations, deletions, and pseudodiploidy, and as increased mutation rate at the hypoxanthine phosphoribosyl transferase (HPRT) gene.

Werner Syndrome has been recognized as an autosomal recessive disorder. Goto et al. (Goto et al., *Nature* 355:735-738, 1992) mapped the WS gene onto the short arm of chromosome 8, using 21 affected Japanese families. The gene is located between marker D8S87 and ankyrin (ANK1). More recently, more refined mapping has pinpointed the WS gene to a region between marker D8S131 and D8S87, an 8.3 cM interval. Identification of the gene and gene product should add considerably to understanding the basis of Werner Syndrome and enable biochemical and genetic approaches to diagnosis and treatment.

The present invention provides a novel, previously unidentified gene for Werner Syndrome and compositions for diagnosis and treatment of WS, and further provides other related advantages.

SUMMARY OF THE INVENTION

Briefly stated, the present invention provides isolated nucleic acid molecules encoding the WRN gene, as well as portions thereof, representative of which are provided in the Figures. The protein which is encoded by the WRN gene is referred to hereinafter as the "WRN protein". Within other embodiments, nucleic acid molecules are provided which encode a mutant WRN gene product that increases the

Within other aspects of the present invention, isolated nucleic acid molecules are provided, selected from the group consisting of (a) an isolated nucleic acid molecule as set forth in the Figures, or complementary sequence thereof, (b) an isolated nucleic acid molecule that specifically hybridizes to the nucleic acid molecule of (a) under conditions of high stringency, and (c) an isolated nucleic acid that encodes a WRN gene product (WRN protein). As utilized herein, it should be understood that a nucleic acid molecule hybridizes "specifically" to an WRN gene (or related sequence) if it hybridizes detectably to such a sequence, but does not significantly or detectably hybridize to the Bloom's Syndrome gene (Ellis et al., *Cell* 83:655-666, 1995).

Within other aspects, expression vectors are provided comprising a promoter operably linked to one of the nucleic acid molecule described above. Representative examples of suitable promoters include tissue-specific promoters, as well as promoters such as the CMV I-E promoter, SV40 early promoter and Mol.V LTR. Within related aspects, viral vectors are provided that are capable of directing the expression of a nucleic acid molecule as described above. Representative examples of such viral vectors include herpes simplex viral vectors, adenoviral vectors, adenovirus-associated viral vectors and retroviral vectors. Also provided are host cells (e.g., human, dog, monkey, rat or mouse cells) which carry the above-described vectors.

Within other aspects of the present invention, isolated proteins or polypeptides are provided comprising a WRN gene product, as well as peptides of greater than 12, 13 or 20 amino acids. Within another embodiment, the protein is a mutant WRN gene product that increases the probability of Werner's Syndrome.

Within yet another aspect of the present invention, methods of treating or preventing Werner's Syndrome are provided (as well as for related diseases which are discussed in more detail below), comprising the step of administering to a patient a vector containing or expressing a nucleic acid molecule as described above, thereby reducing the likelihood or delaying the onset of Werner's Syndrome (or the related

Figure 1 is a schematic diagram of the WRN gene.

Figure 2 is a schematic diagram of the WRN protein structure.

administering to a patient a protein as described above, thereby reducing the likelihood

or delaying the onset of Werner's Syndrome (or a related disease) in the patient. Within certain embodiments, the above methods may be accomplished by *in vivo* administration.

Also provided by the present invention are pharmaceutical compositions comprising a nucleic acid molecule, vector, host cell, protein, or antibody as described above, along with a pharmaceutically acceptable carrier or diluent.

Within other aspects of the present invention, antibodies are provided which specifically bind to an WRN protein or to unique peptides derived therefrom. As utilized herein, it should be understood that an antibody is specific for an WRN protein (or peptide) if it binds detectably, and with a K_d of $10^{-7}M$ or less (e.g., $10^{-8}M$, $10^{-9}M$, etc.), but does not bind detectably (or with an affinity of greater than $10^{-7}M$, (e.g., $10^{-6}M$, $10^{-5}M$, etc.) to an unrelated helicase (e.g., the Bloom's Syndrome gene, *supra*). Also provided are hybridomas which are capable of producing such antibodies.

Within other aspects of the present invention, nucleic acid probes are provided which are capable of specifically hybridizing (as defined below) to an WRN gene under conditions of high stringency. Within one related aspect, such probes comprise at least a portion of the nucleotide sequence shown in the Figures, or its complementary sequence, the probe being capable of specifically hybridizing to a mutant WRN gene under conditions of high stringency. Representative probes of the present invention are generally at least 12 nucleotide bases in length, although they may be 14, 16, 18 bases or longer. Also provided are primer pairs capable of specifically amplifying all or a portion of any of the nucleic acid molecules disclosed herein.

Within other aspects of the invention, methods are provided for diagnosing a patient having an increased likelihood of contracting Werner's Syndrome (or a related disease), comprising the steps of (a) obtaining from a patient a biological sample containing nucleic acid, (b) incubating the nucleic acid with a probe which is capable of specifically hybridizing to a mutant WRN gene under conditions of high stringency, and (c) determining that said patient is at an increased risk for

contracting Werner's Syndrome (or a related disease). Within another aspect,

methods are provided comprising the steps of (a) obtaining from a patient a biological sample containing nucleic acid, (b) amplifying a selected nucleic acid sequence associated with a mutant WRN gene, and (c) detecting the presence of an amplified nucleic acid sequence, and thereby determining that the patient has an increased
5 likelihood of contracting Werner's Syndrome (or a related disease). Suitable biological samples include nucleated cells obtained from the peripheral blood, from buccal swabs, or brain tissue.

Within another aspect, peptide vaccines are provided which comprise a portion of a mutant WRN gene product containing a mutation, in combination with a
10 pharmaceutically acceptable carrier or diluent.

Within yet another aspect, transgenic animals are provided whose germ cells and somatic cells contain a WRN gene (or lack thereof, *i.e.*, a "knockout") which is operably linked to a promoter effective for the expression of the gene, the gene being introduced into the animal, or an ancestor of the animal, at an embryonic stage. Within
15 one embodiment, the animal is a mouse, rat or dog. Within other embodiments, the WRN gene is expressed from a vector as described above. Within yet another embodiment, the WRN gene encodes a mutant WRN gene product.

These and other aspects of the present invention will become evident upon reference to the following detailed description and attached drawings. In addition,
20 various references are set forth herein which describe in more detail certain procedures or compositions (*e.g.*, plasmids, etc.), and are therefore incorporated by reference in their entirety.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE LISTING

25 Figure 1 is a genetic and physical map of the WRN region. The genetic map (A) of the region is sex-equal with distances given in cM. The polymorphic loci used (B) are di-nucleotide and tri-nucleotide repeat STRP loci. The physical map

30 2253, 3833, 2236, and 3101. Marker order was determined from the sequence-tagged

site (STS) content of YACs, P1 clones, and cosmid clones and from genomic DNA sequence from P1 clones. The YACs presented (D) represent the minimal tiling and are the YACs used for cDNA selection experiments. The P1 and cosmid clones needed for the minimum tiling path are shown (E). Clones shown are P1 clones except for 8C11, which is a cosmid clone. Clone order was established by STS content.

Figures 2A and 2B are the DNA (SEQ ID No. 70) and predicted amino acid (SEQ ID No. 71) sequences of the WRN gene transcript. The one-letter amino acid code is used in Fig. 2B.

Figures 3A-3C are the DNA and predicted amino acid sequence of an alternate WRN gene transcript (SEQ ID Nos. 72 and 73).

Figures 4A-4G are an alignment of the WRN gene product (SEQ ID No. 74) with known helicases from *S. pombe* (SEQ ID No. 76), *E. coli* (SEQ ID No. 75), human (SEQ ID No. 77) and the Bloom's Syndrome gene "BLM" (SEQ ID No. 78).

Figures 5A-5U are the genomic DNA sequence of the region containing a WRN gene (SEQ ID No. 79).

Figure 6 presents a cDNA sequence of the mouse WRN gene (SEQ ID Nos. 205 and 206).

Figure 7 is a genomic DNA sequence of the mouse WRN gene (SEQ ID Nos. 207-209).

Figure 8 is a diagram of the WRN gene product with location of mutations. A, WRN cDNA. Numbering across the top refers to the cDNA sequence as numbered in GenBank L76937. B, Predicted WRN gene product. The helicase domain is designated as "HD", motifs from I to VI are indicated. C, Location of mutations. Numbering across the bottom refer to the mutations. *: nonsense mutation. ^: frame shift mutation caused by a single base deletion. Gray lines: frame shift mutations causing deletion of exon(s). D, Predicted proteins. Lines represent the different predicted truncated proteins produced from mutations in the WRN gene.

Figure 9 shows WRN gene product by indirect antibody staining (panel A), indirect (panel B), and the size of cells (panel C) expressing the WRN gene.

Figure 10 shows the alignment of the mouse and human WRN gene products.

DETAILED DESCRIPTION OF THE INVENTION

5 Definitions

Prior to setting forth the invention in detail, it may be helpful to an understanding thereof to set forth definitions of certain terms and to list and to define the abbreviations that will be used hereinafter.

10 "Genetic marker" is any segment of a chromosome that is distinguishably unique in the genome, and polymorphic in the population so as to provide information about the inheritance of linked DNA sequences, genes and/or other markers.

 "Vector" refers to an assembly which is capable of directing the expression of a WRN gene, as well as any additional sequence(s) or gene(s) of interest.
15 The vector must include transcriptional promoter elements which are operably linked to the genes of interest. The vector may be composed of either deoxyribonucleic acids ("DNA"), ribonucleic acids ("RNA"), or a combination of the two (e.g., a DNA-RNA chimeric). Optionally, the vector may include a polyadenylation sequence, one or more restriction sites, as well as one or more selectable markers such as neomycin
20 phosphotransferase or hygromycin phosphotransferase. Additionally, depending on the host cell chosen and the vector employed, other genetic elements such as an origin of replication, additional nucleic acid restriction sites, enhancers, sequences conferring inducibility of transcription, and selectable markers, may also be incorporated into the vectors described herein.

25 Abbreviations: YAC, yeast artificial chromosome; EST, expressed sequence tag; PCR, polymerase chain reaction; RT-PCR, PCR process in which RNA is first transcribed into DNA at the first step using reverse transcriptase (RT); cDNA, any

30 compositions for the detection and treatment of Werner's Syndrome, as well as related

diseases. These methods and compositions include a family of Werner's Syndrome-related genes, and the proteins encoded thereby, that have been implicated in the onset of Werner's Syndrome. These genes and proteins, including genetic markers, nucleic acid sequences and clones, are also useful in the creation of *in vitro* and animal models and screening tests useful for the study of Werner's Syndrome, including the possible identification of other genes implicated in Werner's Syndrome. The present invention also provides vector constructs, genetic markers, nucleic acid sequences, clones, diagnostic tests and compositions and methods for the identification of individuals likely to suffer from Werner's Syndrome.

Genes and Gene Products Related To Werner's Syndrome

The present invention provides isolated nucleic acid molecules comprising a portion of the gene which is implicated in the onset of WS. Briefly, as can be seen from Figure 4, this gene encodes a protein that is similar in amino acid sequence to several known ATP-dependent DNA helicases (enzymes that unwind the DNA duplex). It is less similar to known RNA-DNA helicases. Helicases are involved in the replication of DNA, often binding the replication origin, and/or the replication complex. In addition, the single stranded DNA that is involved in recombination can be generated by DNA helicases.

Although various aspects of the WRN gene (or portions thereof) are shown in the Figures, it should be understood that within the context of the present invention, reference to one or more of these genes includes derivatives of the genes that are substantially similar to the genes (and, where appropriate, the proteins (including peptides and polypeptides) that are encoded by the genes and their derivatives). As used herein, a nucleotide sequence is deemed to be "substantially similar" if: (a) the nucleotide sequence is derived from the coding region of the described genes and includes, for example, portions of the sequence or allelic variations of the sequences

hybridization to nucleotide sequences of the present invention under high or very high

8

10

15

20

25

cDNA libraries may be designed based on the sequence provided herein. Preferably, the oligonucleotide is 20-30 bases long. Such an oligonucleotide may be synthesized by automated synthesis. The oligonucleotide may be conveniently labeled at the 5' end with a reporter molecule, such as a radionuclide, (e.g., ^{32}P) or biotin. The library is
5 plated as colonies or phage, depending upon the vector, and the recombinant DNA is transferred to nylon or nitrocellulose membranes. Following denaturation, neutralization, and fixation of the DNA to the membrane, the membranes are hybridized with the labeled probe. The membranes are washed and the reporter molecule detected. The hybridizing colonies or phage are isolated and propagated. Candidate clones or
10 PCR amplified fragments may be verified as containing WRN DNA by any of various means. For example, the candidate clones may be hybridized with a second, nonoverlapping probe or subjected to DNA sequence analysis. In these ways, clones containing WRN gene, which are suitable for use in the present invention are isolated.

The structure of the proteins encoded by the nucleic acid molecules
15 described herein may be predicted from the primary translation products using the hydrophobicity plot function of, for example, P/C Gene, Lasergen System, DNA STAR, Madison, Wisconsin, or according to the methods described by Kyte and Doolittle (*J. Mol. Biol.* 157:105-132, 1982).

WRN proteins of the present invention may be prepared in the form of
20 acidic or basic salts, or in neutral form. In addition, individual amino acid residues may be modified by oxidation or reduction. Furthermore, various substitutions, deletions, or additions may be made to the amino acid or nucleic acid sequences, the net effect of which is to retain or further enhance or decrease the biological activity of the mutant or wild-type protein. Moreover, due to degeneracy in the genetic code, for example, there
25 may be considerable variation in nucleotide sequences encoding the same amino acid sequence.

Other derivatives of the WRN proteins disclosed herein include

30 which may be added to facilitate purification or identification of WRN proteins (see

U.S. Patent No. 4,851,341; *see also*, Hopp et al., *Bio/Technology* 6:1204, 1988.) Alternatively, fusion proteins such as WRN protein- β -galactosidase or WRN protein-luciferase may be constructed in order to assist in the identification, expression, and analysis of WRN proteins.

5 WRN proteins of the present invention may be constructed using a wide variety of techniques described herein. Further, mutations may be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes a derivative having the desired
10 amino acid insertion, substitution, or deletion.

Alternatively, oligonucleotide-directed site-specific (or segment specific) mutagenesis procedures may be employed to provide an altered gene having particular codons altered according to the substitution, deletion, or insertion required. Exemplary methods of making the alterations set forth above are disclosed by Walder et al. (*Gene* 15 42:133, 1986); Bauer et al. (*Gene* 37:73, 1985); Craik (*BioTechniques*, January 1985, 12-19); Smith et al. (*Genetic Engineering: Principles and Methods*, Plenum Press, 1981); and Sambrook et al. (*supra*). Deletion or truncation derivatives of WRN proteins (*e.g.*, a soluble extracellular portion) may also be constructed by utilizing convenient restriction endonuclease sites adjacent to the desired deletion. Subsequent
20 to restriction, overhangs may be filled in, and the DNA religated. Exemplary methods of making the alterations set forth above are disclosed by Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, 1989).

Mutations of the present invention preferably preserve the reading frame of the coding sequences. Furthermore, the mutations will preferably not create
25 complementary regions that could hybridize to produce secondary mRNA structures, such as loops or hairpins, that would adversely affect translation of the mRNA. Although a mutation site may be predetermined, it is not necessary that the nature of the
target codon and the expressed mutants screened for indicative biological activity

Alternatively, mutations may be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes a derivative having the desired amino acid insertion, substitution, or deletion.

WRN proteins may also be constructed utilizing techniques of PCR mutagenesis, chemical mutagenesis (Drinkwater and Klinedinst, *PNAS* 83:3402-3406, 1986), by forced nucleotide misincorporation (*e.g.*, Liao and Wise *Gene* 88:107-111, 1990), or by use of randomly mutagenized oligonucleotides (Horwitz et al., *Genome* 3:112-117, 1989).

Proteins can be isolated by, among other methods, culturing suitable host and vector systems to produce the recombinant translation products of the present invention. Supernates from such cell lines, or protein inclusions or whole cells where the protein is not excreted into the supernate, can then be treated by a variety of purification procedures in order to isolate the desired proteins. For example, the supernate may be first concentrated using commercially available protein concentration filters, such as an Amicon or Millipore Pellicon ultrafiltration unit. Following concentration, the concentrate may be applied to a suitable purification matrix such as, for example, an anti-protein antibody bound to a suitable support. Alternatively, anion or cation exchange resins may be employed in order to purify the protein. As a further alternative, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps may be employed to further purify the protein. Other methods of isolating the proteins of the present invention are well known in the skill of the art.

A protein is deemed to be "isolated" within the context of the present invention if no other (undesired) protein is detected pursuant to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis followed by Coomassie blue staining. Within other embodiments, the desired protein can be isolated such that no

Expression of a WRN gene

The present invention also provides for the manipulation and expression of the above described genes by culturing host cells containing a vector capable of expressing the above-described genes. Such vectors or vector constructs include either
5 synthetic or cDNA-derived nucleic acid molecules encoding WRN proteins, which are operably linked to suitable transcriptional or translational regulatory elements. Suitable regulatory elements may be derived from a variety of sources, including bacterial, fungal, viral, mammalian, insect, or plant genes. Selection of appropriate regulatory elements is dependent on the host cell chosen, and may be readily accomplished by one
10 of ordinary skill in the art. Examples of regulatory elements include: a transcriptional promoter and enhancer or RNA polymerase binding sequence, a transcriptional terminator, and a ribosomal binding sequence, including a translation initiation signal.

Nucleic acid molecules that encode any of the WRN proteins described above may be readily expressed by a wide variety of prokaryotic and eukaryotic host
15 cells, including bacterial, mammalian, yeast or other fungi, viral, insect, or plant cells. Methods for transforming or transfecting such cells to express foreign DNA are well known in the art (see, e.g., Itakura et al., U.S. Patent No. 4,704,362; Hinnen et al., *Proc. Natl. Acad. Sci. USA* 75:1929-1933, 1978; Murray et al., U.S. Patent No. 4,801,542; Upshall et al., U.S. Patent No. 4,935,349; Hagen et al., U.S. Patent No. 4,784,950; Axel
20 et al., U.S. Patent No. 4,399,216; Goeddel et al., U.S. Patent No. 4,766,075; and Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory Press, 1989; for plant cells see Czako and Marton, *Plant Physiol* 104:1067-1071, 1994; and Paszkowski et al., *Biotech.* 24:387-392, 1992).

Bacterial host cells suitable for carrying out the present invention include
25 *E. coli*, *B. subtilis*, *Salmonella typhimurium*, and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, as well as many other bacterial species well known to one of ordinary skill in the art. Representative examples of

such vectors include: (a) a minimal promoter, (b) a bacterial origin of replication, (c) one or more selectable phenotypic markers, and a bacterial

origin of replication. Representative promoters include the β -lactamase (penicillinase) and lactose promoter system (see Chang et al., *Nature* 275:615, 1978), the T7 RNA polymerase promoter (Studier et al., *Meth. Enzymol.* 185:60-89, 1990), the lambda promoter (Elvin et al., *Gene* 87:123-126, 1990), the *trp* promoter (Nichols and Yanofsky, *Meth. in Enzymology* 101:155, 1983) and the *tac* promoter (Russell et al., *Gene* 20: 231, 1982). Representative selectable markers include various antibiotic resistance markers such as the kanamycin or ampicillin resistance genes. Many plasmids suitable for transforming host cells are well known in the art, including among others, pBR322 (see Bolivar et al., *Gene* 2:95, 1977), the pUC plasmids pUC18, pUC19, pUC118, pUC119 (see Messing, *Meth. in Enzymology* 101:20-77, 1983 and Vieira and Messing, *Gene* 19:259-268, 1982), and pNH8A, pNH16a, pNH18a, and Bluescript M13 (Stratagene, La Jolla, Calif.).

Yeast and fungi host cells suitable for carrying out the present invention include, among others, *Saccharomyces pombe*, *Saccharomyces cerevisiae*, the genera *Pichia* or *Kluyveromyces* and various species of the genus *Aspergillus* (McKnight et al., U.S. Patent No. 4,935,349). Suitable expression vectors for yeast and fungi include, among others, YCp50 (ATCC No. 37419) for yeast, and the amdS cloning vector pV3 (Turnbull, *Bio/Technology* 7:169, 1989), YRp7 (Struhl et al., *Proc. Natl. Acad. Sci. USA* 76:1035-1039, 1978), YEp13 (Broach et al., *Gene* 8:121-133, 1979), pJDB249 and pJDB219 (Beggs, *Nature* 275:104-108, 1978) and derivatives thereof.

Preferred promoters for use in yeast include promoters from yeast glycolytic genes (Hitzeman et al., *J. Biol. Chem.* 255:12073-12080, 1980; Alber and Kawasaki, *J. Mol. Appl. Genet.* 1:419-434, 1982) or alcohol dehydrogenase genes (Young et al., in *Genetic Engineering of Microorganisms for Chemicals*, Hollaender et al. (eds.), p. 355, Plenum, New York, 1982; Ammerer, *Meth. Enzymol.* 101:192-201, 1983). Examples of useful promoters for fungi vectors include those derived from *Aspergillus nidulans* glycolytic genes, such as the *adh3* promoter (McKnight et al.

et al., *et al.*, 1985).

As with bacterial vectors, the yeast vectors will generally include a selectable marker, which may be one of any number of genes that exhibit a dominant phenotype for which a phenotypic assay exists to enable transformants to be selected. Preferred selectable markers are those that complement host cell auxotrophy, provide
5 antibiotic resistance or enable a cell to utilize specific carbon sources, and include *leu2* (Broach et al., *ibid.*), *ura3* (Botstein et al., *Gene* 8:17, 1979), or *his3* (Struhl et al., *ibid.*). Another suitable selectable marker is the *cat* gene, which confers chloramphenicol resistance on yeast cells.

Techniques for transforming fungi are well known in the literature, and
10 have been described, for instance, by Beggs (*ibid.*), Hinnen et al. (*Proc. Natl. Acad. Sci. USA* 75:1929-1933, 1978), Yelton et al. (*Proc. Natl. Acad. Sci. USA* 81:1740-1747, 1984), and Russell (*Nature* 301:167-169, 1983). The genotype of the host cell may contain a genetic defect that is complemented by the selectable marker present on the expression vector. Choice of a particular host and selectable marker is well within the
15 level of ordinary skill in the art.

Protocols for the transformation of yeast are also well known to those of ordinary skill in the art. For example, transformation may be readily accomplished either by preparation of spheroplasts of yeast with DNA (*see* Hinnen et al., *PNAS USA* 75:1929, 1978) or by treatment with alkaline salts such as LiCl (*see* Itoh et al., *J*
20 *Bacteriology* 153:163, 1983). Transformation of fungi may also be carried out using polyethylene glycol as described by Cullen et al. (*Bio/Technology* 5:369, 1987).

Viral vectors include those which comprise a promoter that directs the expression of an isolated nucleic acid molecule that encodes an WRN protein as described above. A wide variety of promoters may be utilized within the context of the
25 present invention, including for example, promoters such as MoMLV LTR, RSV LTR, Friend MuLV LTR, adenoviral promoter (Ohno et al., *Science* 265: 781-784, 1994), neomycin phosphotransferase promoter/enhancer, late parvovirus promoter (Koering

et al., *Proc. Natl. Acad. Sci. USA* 84:1000-1004, 1987), and the cytomegalovirus immediate late promoter. Within particularly

preferred embodiments of the invention, the promoter is a tissue-specific promoter (see e.g., WO 91/02805; EP 0,415,731; and WO 90/07936). Representative examples of suitable tissue specific promoters include neural specific enolase promoter, platelet derived growth factor beta promoter, bone morpho-genetic protein promoter, human 5 alpha1-chimaerin promoter, synapsin I promoter and synapsin II promoter. In addition to the above-noted promoters, other viral-specific promoters (e.g., retroviral promoters (including those noted above, as well as others such as HIV promoters), hepatitis, herpes (e.g., EBV), and bacterial, fungal or parasitic (e.g., malarial) -specific promoters may be utilized in order to target a specific cell or tissue which is infected with a virus, 10 bacteria, fungus or parasite.

Thus, WRN proteins of the present invention may be expressed from a variety of viral vectors, including for example, herpes viral vectors (e.g., U.S. Patent No. 5,288,641), adenoviral vectors (e.g., WO 94/26914, WO 93/9191; Kolls et al., *PNAS* 91(1):215-219, 1994; Kass-Eisler et al., *PNAS* 90(24):11498-502, 1993; Guzman et al., *Circulation* 88(6):2838-48, 1993; Guzman et al., *Cir. Res* 73(6):1202-1207, 15 1993; Zabner et al., *Cell* 75(2):207-216, 1993; Li et al., *Hum Gene Ther.* 4(4):403-409, 1993; Caillaud et al., *Eur. J. Neurosci.* 5(10):1287-1291, 1993; Vincent et al., *Nat. Genet.* 5(2):130-134, 1993; Jaffe et al., *Nat. Genet.* 1(5):372-378, 1992; and Levrero et al., *Gene* 101(2):195-202, 1991), adeno-associated viral vectors (WO 95/13365; Flotte et al., *PNAS* 90(22):10613-10617, 1993), baculovirus vectors, parvovirus vectors (Koering et al., *Hum. Gene Therap.* 5:457-463, 1994), pox virus vectors (Panicali and Paoletti, *PNAS* 79:4927-4931, 1982; and Ozaki et al., *Biochem. Biophys. Res. Comm.* 193(2):653-660, 1993), and retroviruses (e.g., EP 0,415,731; WO 90/07936; WO 91/0285, WO 94/03622; WO 93/25698; WO 93/25234; U.S. Patent No. 5,219,740; WO 20 93/11230; WO 93/10218. Viral vectors may likewise be constructed which contain a mixture of different elements (e.g., promoters, envelope sequences and the like) from different viruses, or non-viral sources. Within various embodiments, either the viral

Mammalian cells suitable for carrying out the present invention include, among others: PC12 (ATCC No. CRL1721), N1E-115 neuroblastoma, SK-N-BE(2)C neuroblastoma, SHSY5 adrenergic neuroblastoma, NS20Y and NG108-15 murine cholinergic cell lines, or rat F2 dorsal root ganglion line, COS (*e.g.*, ATCC No. CRL 1650 or 1651), BHK (*e.g.*, ATCC No. CRL 6281; BHK 570 cell line (deposited with the American Type Culture Collection under accession number CRL 10314), CHO (ATCC No. CCL 61), HeLa (*e.g.*, ATCC No. CCL 2), 293 (ATCC No. 1573; Graham et al., *J. Gen. Virol.* 36:59-72, 1977) and NS-1 cells. Other mammalian cell lines may be used within the present invention, including Rat Hep I (ATCC No. CRL 1600), Rat Hep II (ATCC No. CRL 1548), TCMK (ATCC No. CCL 139), Human lung (ATCC No. CCL 75.1), Human hepatoma (ATCC No. HTB-52), Hep G2 (ATCC No. HB 8065), Mouse liver (ATCC No. CCL 29.1), NCTC 1469 (ATCC No. CCL 9.1), SP2/0-Ag14 (ATCC No. 1581), HIT-T15 (ATCC No. CRL 1777), and RINm 5AHT₂B (Orskov and Nielson, *FEBS* 229(1):175-178, 1988).

Mammalian expression vectors for use in carrying out the present invention will include a promoter capable of directing the transcription of a cloned gene or cDNA. Preferred promoters include viral promoters and cellular promoters. Viral promoters include the cytomegalovirus immediate early promoter (Boshart et al., *Cell* 41:521-530, 1985), cytomegalovirus immediate late promoter, SV40 promoter (Subramani et al., *Mol. Cell. Biol.* 1:854-864, 1981), MMTV LTR, RSV LTR, metallothionein-1, adenovirus E1a. Cellular promoters include the mouse metallothionein-1 promoter (Palmiter et al., U.S. Patent No. 4,579,821), a mouse V_K promoter (Bergman et al., *Proc. Natl. Acad. Sci. USA* 81:7041-7045, 1983; Grant et al., *Nucl. Acids Res.* 15:5496, 1987) and a mouse V_H promoter (Loh et al., *Cell* 33:85-93, 1983). The choice of promoter will depend, at least in part, upon the level of expression desired or the recipient cell line to be transfected.

Such expression vectors may also contain a set of RNA splice sites

adenovirus and/or immunoglobulin genes. Also contained in the expression vectors is a

polyadenylation signal located downstream of the coding sequence of interest. Suitable polyadenylation signals include the early or late polyadenylation signals from SV40 (Kaufman and Sharp, *ibid.*), the polyadenylation signal from the Adenovirus 5 E1B region and the human growth hormone gene terminator (DeNoto et al., *Nuc. Acids Res.* 9:3719-3730, 1981). The expression vectors may include a noncoding viral leader sequence, such as the Adenovirus 2 tripartite leader, located between the promoter and the RNA splice sites. Preferred vectors may also include enhancer sequences, such as the SV40 enhancer. Expression vectors may also include sequences encoding the adenovirus VA RNAs. Suitable expression vectors can be obtained from commercial sources (e.g., Stratagene, La Jolla, Calif.).

Vector constructs comprising cloned DNA sequences can be introduced into cultured mammalian cells by, for example, calcium phosphate-mediated transfection (Wigler et al., *Cell* 14:725, 1978; Corsaro and Pearson, *Somatic Cell Genetics* 7:603, 1981; Graham and Van der Eb, *Virology* 52:456, 1973), electroporation (Neumann et al., *EMBO J.* 1:841-845, 1982), or DEAE-dextran mediated transfection (Ausubel et al. (eds.), *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc., NY, 1987). To identify cells that have stably integrated the cloned DNA, a selectable marker is generally introduced into the cells along with the gene or cDNA of interest. Preferred selectable markers for use in cultured mammalian cells include genes that confer resistance to drugs, such as neomycin, hygromycin, and methotrexate. The selectable marker may be an amplifiable selectable marker. Preferred amplifiable selectable markers are the DHFR gene and the neomycin resistance gene. Selectable markers are reviewed by Thilly (*Mammalian Cell Technology*, Butterworth Publishers, Stoneham, MA, which is incorporated herein by reference).

Mammalian cells containing a suitable vector are allowed to grow for a period of time, typically 1-2 days, to begin expressing the DNA sequence(s) of interest. Drug selection is then applied to select for growth of cells that are expressing the

desired selectable marker. In drug concentrations that are increased in a stepwise manner to select for increased copy number of the cloned sequences, thereby increasing

expression levels. Cells expressing the introduced sequences are selected and screened for production of the protein of interest in the desired form or at the desired level. Cells that satisfy these criteria can then be cloned and scaled up for production.

5 Protocols for the transfection of mammalian cells are well known to those of ordinary skill in the art. Representative methods include calcium phosphate mediated transfection, electroporation, lipofection, retroviral, adenoviral and protoplast fusion-mediated transfection (*see* Sambrook et al., *supra*). Naked vector constructs can also be taken up by muscular cells or other suitable cells subsequent to injection into the muscle of a mammal (or other animals).

10 Numerous insect host cells known in the art can also be useful within the present invention, in light of the subject specification. For example, the use of baculoviruses as vectors for expressing heterologous DNA sequences in insect cells has been reviewed by Atkinson et al. (*Pestic. Sci.* 28:215-224,1990).

15 Numerous plant host cells known in the art can also be useful within the present invention, in light of the subject specification. For example, the use of *Agrobacterium rhizogenes* as vectors for expressing genes in plant cells has been reviewed by Sinkar et al., (*J. Biosci. (Bangalore)* 11:47-58, 1987).

20 WRN proteins may be prepared by growing (typically by culturing) the host/vector systems described above, in order to express the recombinant WRN proteins. Recombinantly produced WRN proteins may be further purified as described in more detail below.

25 Within related aspects of the present invention, WRN proteins may be expressed in a transgenic animal whose germ cells and somatic cells contain a WRN gene which is operably linked to a promoter effective for the expression of the gene. Alternatively, in a similar manner transgenic animals may be prepared that lack the WRN gene (*e.g.*, "knockout" mice). Such transgenics may be prepared in a variety non-human animals, including mice, rats, rabbits, sheep, dogs, goats and pigs (*see* Hammer

41:343-345, 1985 and U.S. Patent Nos. 5,175,383, 5,087,571, 4,736,866, 5,387,742, 5,347,075, 5,221,778, and 5,175,384).

Briefly, an expression vector, including a nucleic acid molecule to be expressed together with appropriately positioned expression control sequences, is introduced into pronuclei of fertilized eggs, for example, by microinjection. Integration of the injected DNA is detected by blot analysis of DNA from tissue samples. It is preferred that the introduced DNA be incorporated into the germ line of the animal so that it is passed on to the animal's progeny. Tissue-specific expression may be achieved through the use of a tissue-specific promoter, or through the use of an inducible promoter, such as the metallothionein gene promoter (Palmiter et al., 1983, *ibid*), which allows regulated expression of the transgene.

Vectors of the present invention may contain or express a wide variety of additional nucleic acid molecules in place of or in addition to an WRN protein as described above, either from one or several separate promoters. For example, the viral vector may express a lymphokine or lymphokine receptor, antisense or ribozyme sequence or toxins. Representative examples of lymphokines include IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, GM-CSF, G-CSF, M-CSF, alpha-interferon, beta-interferon, gamma-interferon, and tumor necrosis factors, as well as their respective receptors. Representative examples of antisense sequences include antisense sequences which block the expression of WRN protein mutants. Representative examples of toxins include: ricin, abrin, diphtheria toxin, cholera toxin, saporin, gelonin, pokeweed antiviral protein, tritin, *Shigella* toxin, and *Pseudomonas* exotoxin A.

Within other aspects of the invention, antisense oligonucleotide molecules are provided which specifically inhibit expression of mutant WRN nucleic acid sequences (*see generally*, Hirashima et al. in *Molecular Biology of RNA: New Perspectives* (M. Inouye and B. S. Dudock, eds., 1987 Academic Press, San Diego, p.

198-199), U.S. 5,359,951; WO 92/06693; and EP-A2-612844). Briefly, such

molecules are constructed such that they are complementary to, and able to form Watson-Crick base pairs with, a region of transcribed WRN mutant mRNA sequence containing an WRN mutation. The resultant double-stranded nucleic acid interferes with subsequent processing of the mRNA, thereby preventing protein synthesis.

Within other related aspects of the invention, ribozyme molecules are provided wherein an antisense oligonucleotide sequence is incorporated into a ribozyme which can specifically cleave mRNA molecules transcribed from a mutant WRN gene (see generally, Kim et al. *Proc. Nat. Acad. Sci. USA* 84:8788 (1987); Haseloff, et al. *Nature* 234:585 (1988), Cech, *JAMA* 260:3030 (1988); Jeffries, et al. *Nucleic Acids Res.* 17:1371 (1989); U.S. 5,093,246; U.S. 5,354,855; U.S. 5,144,019; U.S. 5,272,262; U.S. 5,254,678; and U.S. 4,987,071). According to this aspect of the invention, the antisense sequence which is incorporated into a ribozyme includes a sequence complementary to, and able to form Watson-Crick base pairs with, a region of the transcribed mutant WRN mRNA containing an WRN mutation. The antisense sequence thus becomes a targeting agent for delivery of catalytic ribozyme activity specifically to mutant WRN mRNA, where such catalytic activity cleaves the mRNA to render it incapable of being subsequently processed for WRN protein translation.

Host Cells

As discussed above, nucleic acid molecules which encode the WRN proteins of the present invention (or the vectors which contain and/or express related mutants) may readily be introduced into a wide variety of host cells. Representative examples of such host cells include plant cells, eukaryotic cells, and prokaryotic cells. Within preferred embodiments, the nucleic acid molecules are introduced into cells from a vertebrate or warm-blooded animal, such as a human, macaque, dog, cow, horse, pig, sheep, rat, hamster, mouse or fish cell, or any hybrid thereof.

Preferred prokaryotic host cells for use within the present invention include:

(a) *Escherichia coli*,
and
(b) *Bacillus subtilis*.

Sequences cloned therein are well known in the art (see, e.g., Maniatis et al., *Molecular*

Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, 1982, which is incorporated herein by reference; or Sambrook et al., *supra*). Vectors used for expressing cloned DNA sequences in bacterial hosts will generally contain a selectable marker, such as a gene for antibiotic resistance, and a promoter that functions in the host cell. Appropriate promoters include the *trp* (Nichols and Yanofsky, *Meth. Enzymol.* 101:155-164, 1983), *lac* (Casadaban et al., *J. Bacteriol.* 143:971-980, 1980), and phage λ (Queen, *J. Mol. Appl. Genet.* 2:1-10, 1983) promoter systems. Plasmids useful for transforming bacteria include the pUC plasmids (Messing, *Meth. Enzymol.* 101:20-78, 1983; Vieira and Messing, *Gene* 19:259-268, 1982), pBR322 (Bolivar et al., *Gene* 2:95-113, 1977), pCQV2 (Queen, *ibid.*), and derivatives thereof. Plasmids may contain both viral and bacterial elements.

Preferred eukaryotic cells include cultured mammalian cell lines (e.g., rodent or human cell lines) and fungal cells, including species of yeast (e.g., *Saccharomyces* spp., particularly *S. cerevisiae*, *Schizosaccharomyces* spp., or *Kluyveromyces* spp.) or filamentous fungi (e.g., *Aspergillus* spp., *Neurospora* spp.). Strains of the yeast *Saccharomyces cerevisiae* are particularly preferred. Methods for producing recombinant proteins in a variety of prokaryotic and eukaryotic host cells are generally known in the art (see, "Gene Expression Technology," *Methods in Enzymology*, Vol. 185, Goeddel (ed.), Academic Press, San Diego, Calif., 1990; see also, "Guide to Yeast Genetics and Molecular Biology," *Methods in Enzymology*, Guthrie and Fink (eds.), Academic Press, San Diego, Calif., 1991). In general, a host cell will be selected on the basis of its ability to produce the protein of interest at a high level or its ability to carry out at least some of the processing steps necessary for the biological activity of the protein. In this way, the number of cloned DNA sequences that must be introduced into the host cell can be minimized and overall yield of biologically active protein can be maximized.

The nucleic acid molecules (or vectors) may be introduced into host cells

by methods known in the art, including transformation, electroporation, and infection. See, e.g., Pearson, *Somatic Cell Genet.* 7:603, 1981; Graham and Van der Eb, *Virology* 52:456,

1973), electroporation (Neumann et al., *EMBO J.* 1:841-845, 1982), retroviral, adenoviral, protoplast fusion-mediated transfection or DEAE-dextran mediated transfection (Ausubel et al., (eds.), *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc., NY, NY, 1987).

5 Host cells containing vector constructs of the present invention are then cultured to express a DNA molecule as described above. The cells are cultured according to standard methods in a culture medium containing nutrients required for growth of the chosen host cells. A variety of suitable media are known in the art and generally include a carbon source, a nitrogen source, essential amino acids, vitamins
10 and minerals, as well as other components, e.g., growth factors or serum, that may be required by the particular host cells. The growth medium will generally select for cells containing the DNA construct(s) by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker on the DNA construct or co-transfected with the DNA construct.

15 Suitable growth conditions for yeast cells, for example, include culturing in a chemically defined medium, comprising a nitrogen source, which may be a non-amino acid nitrogen source or a yeast extract, inorganic salts, vitamins and essential amino acid supplements at a temperature between 4°C and 37°C, with 30°C being particularly preferred. The pH of the medium is preferably maintained at a pH greater
20 than 2 and less than 8, more preferably pH 5-6. Methods for maintaining a stable pH include buffering and constant pH control. Preferred agents for pH control include sodium hydroxide. Preferred buffering agents include succinic acid and Bis-Tris (Sigma Chemical Co., St. Louis, Mo.). Due to the tendency of yeast host cells to hyperglycosylate heterologous proteins, it may be preferable to express the nucleic acid
25 molecules of the present invention in yeast cells having a defect in a gene required for asparagine-linked glycosylation. Such cells are preferably grown in a medium containing an osmotic stabilizer. A preferred osmotic stabilizer is sorbitol

Cultured mammalian cells are generally cultured in commercially available serum-containing or serum-free media. Selection of a medium and growth conditions appropriate for the particular cell line used is well within the level of ordinary skill in the art.

5

Antibodies

Antibodies to the WRN proteins discussed above may readily be prepared given the disclosure provided herein. Such antibodies may, within certain embodiments, specifically recognize wild type WRN protein rather than a mutant WRN protein, mutant WRN protein rather than wild type WRN protein, or equally recognize both the mutant and wild-type forms of WRN protein. Antibodies may be used for isolation of the protein, establishing intracellular localization of the WRN protein, inhibiting activity of the protein (antagonist), or enhancing activity of the protein (agonist). Knowledge of the intracellular location of the WRN gene product may be abnormal in patients with WRN mutations, thus allowing the development of a rapid screening assay. As well, assays for small molecules that interact with the WRN gene product will be facilitated by the development of antibodies and localization studies.

Within the context of the present invention, antibodies are understood to include monoclonal antibodies, polyclonal antibodies, anti-idiotypic antibodies, antibody fragments (e.g., Fab, and F(ab')₂, F_v variable regions, or complementarity determining regions). As discussed above, antibodies are understood to be specific against an WRN protein if it binds with a K_d of greater than or equal to 10⁻⁷M, preferably greater than or equal to 10⁻⁸M. The affinity of a monoclonal antibody or binding partner can be readily determined by one of ordinary skill in the art (see Scatchard, *Ann. N.Y. Acad. Sci.* 51:660-672, 1949).

Briefly, polyclonal antibodies may be readily generated by one of ordinary skill in the art from a variety of warm-blooded animals such as horses, cows,

etc. For example, proteins or peptides, preferably conjugated to a carrier protein, may be cross-linked (with glutaraldehyde) is utilized to immunize the animal, through

30

intraperitoneal, intramuscular, intraocular, or subcutaneous injections, an adjuvant such as Freund's complete or incomplete adjuvant. Merely as an example, a peptide corresponding to residues 1375 through 1387 of the WRN polypeptide sequence is used to raise a rabbit polyclonal antiserum. Following several booster immunizations, samples of serum are collected and tested for reactivity to the WRN protein or peptide. Particularly preferred polyclonal antisera will give a signal on one of these assays that is at least three times greater than background. Once the titer of the animal has reached a plateau in terms of its reactivity to the protein, larger quantities of antisera may be readily obtained either by weekly bleedings, or by exsanguinating the animal.

Monoclonal antibodies may also be readily generated using conventional techniques (see U.S. Patent Nos. RE 32,011, 4,902,614, 4,543,439, and 4,411,993 which are incorporated herein by reference; see also *Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analyses*, Plenum Press, Kennett, McKearn, and Bechtol (eds.), 1980, and *Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988, which are also incorporated herein by reference).

Briefly, within one embodiment a subject animal such as a rat or mouse is injected with an WRN protein or portion thereof as described above. The protein may be admixed with an adjuvant such as Freund's complete or incomplete adjuvant in order to increase the resultant immune response. Between one and three weeks after the initial immunization the animal may be reimmunized with another booster immunization, and tested for reactivity to the protein utilizing assays described above. Once the animal has reached a plateau in its reactivity to the injected protein, it is sacrificed, and organs which contain large numbers of B cells such as the spleen and lymph nodes are harvested.

Cells which are obtained from the immunized animal may be immortalized by transfection with a virus such as the Epstein-Barr virus (EBV) (see

suitable myeloma cell in order to create a 'hybridoma' which secretes monoclonal

antibody. Suitable myeloma lines include, for example, NS-1 (ATCC No. TIB 18), and P3X63 - Ag 8.653 (ATCC No. CRL 1580).

Following the fusion, the cells may be placed into culture plates containing a suitable medium, such as RPMI 1640, or DMEM (Dulbecco's Modified Eagles Medium) (JRH Biosciences, Lenexa, Kansas), as well as additional ingredients, such as fetal bovine serum (FBS, *i.e.*, from Hyclone, Logan, Utah, or JRH Biosciences). Additionally, the medium should contain a reagent which selectively allows for the growth of fused spleen and myeloma cells such as HAT (hypoxanthine, aminopterin, and thymidine) (Sigma Chemical Co., St. Louis, Missouri). After about seven days, the resulting fused cells or hybridomas may be screened in order to determine the presence of antibodies which are reactive against an WRN protein. A wide variety of assays may be utilized to determine the presence of antibodies which are reactive against the proteins of the present invention, including for example countercurrent immuno-electrophoresis, radioimmunoassays, radioimmunoprecipitations, enzyme-linked immuno-sorbent assays (ELISA), dot blot assays, western blots, immunoprecipitation, Inhibition or Competition Assays, and sandwich assays (*see* U.S. Patent Nos. 4,376,110 and 4,486,530; *see also* *Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988). Following several clonal dilutions and reassays, a hybridoma producing antibodies reactive against the WRN protein may be isolated.

Other techniques may also be utilized to construct monoclonal antibodies (*see* William D. Huse et al., "Generation of a Large Combinational Library of the Immunoglobulin Repertoire in Phage Lambda," *Science* 246:1275-1281, December 1989; *see also* L. Sastry et al., "Cloning of the Immunological Repertoire in *Escherichia coli* for Generation of Monoclonal Catalytic Antibodies: Construction of a Heavy Chain Variable Region-Specific cDNA Library," *Proc. Natl. Acad. Sci. USA* 86:5728-5732, August 1989; *see also* Michelle Alting-Mees et al., "Monoclonal Antibody Expression

La Jolla, California, which enables the production of antibodies through recombinant

techniques). Briefly, mRNA is isolated from a B cell population, and utilized to create heavy and light chain immunoglobulin cDNA expression libraries in the λ ImmunoZap(H) and λ ImmunoZap(L) vectors. These vectors may be screened individually or co-expressed to form Fab fragments or antibodies (*see* Huse et al.,
5 *supra*; *see also* Sastry et al., *supra*). Positive plaques may subsequently be converted to a non-lytic plasmid which allows high level expression of monoclonal antibody fragments from *E. coli*.

Similarly, portions or fragments, such as Fab and Fv fragments, of antibodies may also be constructed utilizing conventional enzymatic digestion or
10 recombinant DNA techniques to incorporate the variable regions of a gene which encodes a specifically binding antibody. Within one embodiment, the genes which encode the variable region from a hybridoma producing a monoclonal antibody of interest are amplified using nucleotide primers for the variable region. These primers
15 may be synthesized by one of ordinary skill in the art, or may be purchased from commercially available sources. Stratacyte (La Jolla, Calif.) sells primers for mouse and human variable regions including, among others, primers for $V_{H\alpha}$, $V_{H\beta}$, $V_{H\gamma}$, $V_{H\delta}$, C_{H1} , V_L and C_L regions. These primers may be utilized to amplify heavy or light chain variable regions, which may then be inserted into vectors such as ImmunoZAP™ H or
20 ImmunoZAP™ L (Stratacyte), respectively. These vectors may then be introduced into *E. coli*, yeast, or mammalian-based systems for expression. Utilizing these techniques, large amounts of a single-chain protein containing a fusion of the V_H and V_L domains may be produced (*see* Bird et al., *Science* 242:423-426, 1988). In addition, such techniques may be utilized to change a "murine" antibody to a "human" antibody, without altering the binding specificity of the antibody.

25 Once suitable antibodies have been obtained, they may be isolated or purified by many techniques well known to those of ordinary skill in the art (*see Antibodies: A Laboratory Manual*, Harlow and Lane (eds), Cold Spring Harbor
30 combination of these techniques.

Assays

Assays useful within the context of the present invention include those assays for detecting agonists or antagonists of WRN protein activity. Other assays are useful for the screening of peptide or organic molecule libraries. Still other assays are useful for the identification and/or isolation of nucleic acid molecules and/or peptides within the present invention, the identification of proteins that interact or bind the WRN protein, for diagnosis of a patient with an increased likelihood of contracting Werner's Syndrome, or for diagnosis of a patient with susceptibility to or manifestation of a WRN-related disease.

Nucleic Acid Based Diagnostic Tests

Briefly, another aspect of the present invention provides probes and primers for detecting the WRN genes and/or mutants thereof. In one embodiment of this aspect, probes are provided that are capable of specifically hybridizing to DNA or RNA of the WRN genes. For purposes of the present invention, probes are "capable of hybridizing" to DNA or RNA of the WRN gene if they hybridize to an WRN gene under conditions of either high or moderate stringency (*see* Sambrook et al., *supra*) but not significantly or detectably to the an unrelated helicase gene such as the Bloom's Syndrome gene (Ellis et al., *Cell* 83:655-666, 1995). Preferably, the probe hybridizes to suitable nucleotide sequences under high stringency conditions, such as hybridization in 5x SSPE, 1x Denhardt's solution, 0.1% SDS at 65°C, and at least one wash to remove unhybridized probe in the presence of 0.2x SSC, 1x Denhardt's solution, 0.1% SDS at 65°C. Except as otherwise provided herein, probe sequences are designed to allow hybridization to WRN genes, but not to DNA or RNA sequences from other genes. The probes are used, for example, to hybridize to nucleic acid that is present in a biological sample isolated from a patient. The hybridized probe is then detected, thereby detecting the presence of the WRN gene in the sample. Alternatively, the WRN gene may be amplified and the amplified

product subjected to DNA sequencing. Mutants of WRN may be detected by DNA sequence analysis or hybridization with allele-specific oligonucleotide probes under conditions and for time sufficient to allow hybridization to the specific allele. Typically, the hybridization buffer and wash will contain tetramethyl ammonium chloride or the like (*see* Sambrook et al., *supra*).

Nucleic acid probes of the present invention may be composed of either deoxyribonucleic acids (DNA), ribonucleic acids (RNA), nucleic acid analogues (*e.g.*, peptide nucleic acids), or any combination thereof, and may be as few as about 12 nucleotides in length, usually about 14 to 18 nucleotides in length, and possibly as large as the entire sequence of a WRN gene. Selection of probe size is somewhat dependent upon the use of the probe, and is within the skill of the art.

Suitable probes can be constructed and labeled using techniques that are well known in the art. Shorter probes of, for example, 12 bases can be generated synthetically and labeled with ^{32}P using T₄ polynucleotide kinase. Longer probes of about 75 bases to less than 1.5 kb are preferably generated by, for example, PCR amplification in the presence of labeled precursors such as [α - ^{32}P]dCTP, digoxigenin-dUTP, or biotin-dATP. Probes of more than 1.5 kb are generally most easily amplified by transfecting a cell with a plasmid containing the relevant probe, growing the transfected cell into large quantities, and purifying the relevant sequence from the transfected cells. (*See* Sambrook et al., *supra*.)

Probes can be labeled by a variety of markers, including for example, radioactive markers, fluorescent markers, enzymatic markers, and chromogenic markers. The use of ^{32}P is particularly preferred for marking or labeling a particular probe.

It is a feature of this aspect of the invention that the probes can be utilized to detect the presence of WRN mRNA or DNA within a sample. However, if the relevant sample is present in only a limited number, then it may be beneficial to

sequence, including, for example, RNA amplification (*see* Lizardi et al.,

Bio/Technology 6:1197-1202, 1988; Kramer et al., *Nature* 339:401-402, 1989; Lomeli et al., *Clinical Chem.* 35(9):1826-1831, 1989; U.S. Patent No. 4,786,600), and DNA amplification utilizing LCR or polymerase chain reaction ("PCR") (see, U.S. Patent Nos. 4,683,195, 4,683,202, and 4,800,159) (see also U.S. Patent Nos. 4,876,187 and 5,011,769, which describe an alternative detection/amplification system comprising the use of scissile linkages), or other nucleic acid amplification procedures that are well within the level of ordinary skill in the art. With respect to PCR, for example, the method may be modified as known in the art. Transcriptional enhancement of PCR may be accomplished by incorporation of bacteriophage T7 RNA polymerase promoter sequences in one of the primary oligonucleotides, and immunoenzymatic detection of the products from the enhanced emitter may be effected using anti-RNA:DNA antibodies (Blais, *Appl. Environ. Microbiol.* 60:348-352, 1994). PCR may also be used in combination with reverse dot-blot hybridization (Iida et al., *FEMS Microbiol. Lett.* 114:167-172, 1993). PCR products may be quantitatively analyzed by incorporation of dUTP (Dupl  a et al., *Anal. Biochem.* 212:229-236, 1993), and samples may be filter sampled for PCR-gene probe detection (Bej et al., *Appl. Environ. Microbiol.* 57:3529-3534, 1991).

Within a particularly preferred embodiment, PCR amplification is utilized to detect the WRN DNA. Briefly, as described in greater detail below, a DNA sample is denatured at 95  C in order to generate single-stranded DNA. The DNA sample may be a cDNA generated from RNA. Specific primers are then annealed to the single-stranded DNA at 37  C to 70  C, depending on the proportion of AT/GC in the primers. The primers are extended at 72  C with *Taq* DNA polymerase or other thermostable DNA polymerase in order to generate the opposite strand to the template. These steps constitute one cycle, which may be repeated in order to amplify the selected sequence. For greater specificity, nested PCR may be performed. In nested PCR, a second amplification is performed using a second set of primers derived from sequences

convenient size for determining their sequence. In a preferred embodiment, nested PCR is performed.

Within an alternative preferred embodiment, LCR amplification is utilized for amplification. LCR primers are synthesized such that the 5' base of the upstream primer is capable of hybridizing to a unique base pair in a desired gene to specifically detect an WRN gene.

Within another preferred embodiment, the probes are used in an automated, non-isotopic strategy wherein target nucleic acid sequences are amplified by PCR, and then desired products are determined by a colorimetric oligonucleotide ligation assay (OLA) (Nickerson et al., *Proc. Natl. Acad. Sci. USA* 81:8923-8927, 1990).

Primers for the amplification of a selected sequence should be selected from sequences that are highly specific to WRN (and not, e.g., the Bloom's Syndrome gene, *supra*) and form stable duplexes with the target sequence. The primers should also be non-complementary, especially at the 3' end, should not form dimers with themselves or other primers, and should not form secondary structures or duplexes with other regions of DNA. In general, primers of about 18 to 20 nucleotides are preferred, and can be easily synthesized using techniques well known in the art. PCR products, and other nucleic acid amplification products, may be quantitated using techniques known in the art (Dupl   et al., *Anal. Biochem.* 212:229-236, 1993; Higuchi et al., *Bio/Technology* 11:1026-1030).

Within one embodiment of the invention, nucleic acid diagnostics may be developed which are capable of detecting the presence of Werner's Syndrome, or of various related diseases that may be caused by Werner's Syndrome. Briefly, severe mutations in the WRN gene may lead to Werner's Syndrome, as well as a host of related diseases, including for example, increased frequency of some benign and malignant neoplasms (especially sarcomas), cataracts, cardiovascular disease, osteoporosis, type I

diabetes mellitus.

In addition, many of the related diseases may be associated with mutations in

the WRN gene. For example, diabetes and osteoporosis are often associated with aging. Aging population and individuals with these (or other) diseases are screened for mutations in WRN. Any of the assays described herein may be used. RT-PCR is especially preferred in conjunction with DNA sequence determination. To correlate a mutation or polymorphism with disease, sibling pairs in which one sibling has disease are preferred subjects. Once a mutation is identified, other convenient screening assays may be used to assay particular nucleotide changes.

Since the sequences of the two copies of the gene from non-Werner's affected individuals can be correlated with the medical histories of these patients to define these correspondences, these alleles can therefore be used as diagnostics for susceptibilities to these diseases, once the relationship is defined. Certain non-null forms of the gene, for example, in either the homozygous or heterozygous state may significantly affect the propensity for the carriers to develop, for example, cancer. These propensities can be ascertained by examining the sequences of the gene (both copies) in a statistically significant sample of cancer patients. Other diseases (see above) can be similarly examined for significant correlations with certain alleles. To detect such a causal relationship one can use a chi-squared test, or other statistical test, to examine the significance of any correlation between the appropriate genotypes and the disease state as recorded in the medical records, using standard good practices of medical epidemiology. The sequences that define each of the alleles are then valuable diagnostic indicators for an increased susceptibility to the disease. Thus, from the nucleic acid sequences provided herein, a wide variety of Werner's Syndrome-related diseases may be readily detected.

Another cellular phenotype of the cells from Werner's patients is the increased frequency of deletion mutation in these cells. Clearly, the defective helicase in these cells leads to a specific mutator phenotype, while not rendering the cells hypersensitive to a variety of chemical or physical mutagens that damage DNA, like

some alleles may therefore be diagnostic of this class of medical conditions.

Assays for agonists and antagonists

Also provided by the present invention are agonists or antagonists of the WRN gene product comprising a protein, peptide, chemical, or peptidomimetic that binds to the WRN gene product or interacts with a protein that binds to the WRN gene product such that the binding of the agonist or antagonist affects the activity of the WRN gene product. An agonist will activate or increase the activity of the WRN gene product. An antagonist will inhibit or decrease the activity of the WRN gene product. The activity of the WRN gene product may be measured in an assay, such as a helicase assay or other assay that measures an activity of the WRN gene product. Other assays measure the binding of protein that interacts with WRN and is necessary for its activity.

Agonists and antagonists of the WRN gene product may be used to enhance activity or inhibit activity of the gene product. Such agonists and antagonists may be identified by a variety of methods. For example, proteins that bind and activate WRN may be identified using a yeast 2 hybrid detection system. In this system, the WRN gene is fused to either a DNA-binding domain or an activating domain of a yeast gene such as GAL4. A cDNA library is constructed in a vector such that the inserts are fused to one of the domains. The vectors are co-transfected into yeast and selected for transcriptional activation of a reporter gene (Fields and Song, *Nature* 340: 245, 1989). The protein(s) that bind to WRN are candidate agonists. Three different proteins that bind WRN have been identified in an initial screen using the 2-hybrid system.

When the binding site on WRN gene product is determined, molecules that bind and activate WRN protein can be designed and evaluated. For example, computer modeling of the binding site can be generated and mimetics that bind can be designed. Antibodies to the binding site may be generated and analogues of native binding proteins generated as well. Any of these molecules is tested for agonist or antagonist activity by a functional assay of the WRN gene product. For example, to test

administered and activation is monitored. An antagonist will inhibit the activation of

the reporter gene by at least 50%. Similarly, agonist activity may be measured by either enhancing WRN activity in a yeast 2-hybrid system or by coupling the test compound to a DNA binding or activation domain and monitoring activity of the reporter gene.

Labels

WRN proteins, nucleic acid molecules which encodes such proteins, anti-WRN protein antibodies and agonists or antagonists, as described above and below, may be labeled with a variety of molecules, including for example, fluorescent molecules, toxins, and radionuclides. Representative examples of fluorescent molecules include fluorescein, *Phycobili* proteins, such as phycoerythrin, rhodamine, Texas red and luciferase. Representative examples of toxins include ricin, abrin diphtheria toxin, cholera toxin, gelonin, pokeweed antiviral protein, tritin, *Shigella* toxin, and *Pseudomonas* exotoxin A. Representative examples of radionuclides include Cu-64, Ga-67, Ga-68, Zr-89, Ru-97, Tc-99m, Rh-105, Pd-109, In-111, I-123, I-125, I-131, Re-186, Re-188, Au-198, Au-199, Pb-203, At-211, Pb-212 and Bi-212. In addition, the antibodies described above may also be labeled or conjugated to one partner of a ligand binding pair. Representative examples include avidin-biotin, and riboflavin-riboflavin binding protein.

Methods for conjugating or labeling the WRN proteins, nucleic acid molecules which encode such proteins, anti-WRN protein antibodies and agonists or antagonists, as discussed above, with the representative labels set forth above may be readily accomplished by one of ordinary skill in the art (see Trichothecene Antibody Conjugate, U.S. Patent No. 4,744,981; Antibody Conjugate, U.S. Patent No. 5,106,951; Fluorogenic Materials and Labeling Techniques, U.S. Patent No. 4,018,884; Metal Radionuclide Labeled Proteins for Diagnosis and Therapy, U.S. Patent No. 4,897,255; and Metal Radionuclide Chelating Compounds for Improved Chelation Kinetics, U.S. Patent No. 4,988,496; see also Inman, *Methods In Enzymology*, Vol. 34, "Affinity

Bioanalytical Applications," *Anal. Biochem.* 171:1-32, 1988).

Pharmaceutical Compositions

As noted above, the present invention also provides a variety of pharmaceutical compositions, comprising one of the above-described WRN proteins, nucleic acid molecules, vectors, antibodies, host cells, agonists or antagonists, along with a pharmaceutically or physiologically acceptable carrier, excipients or diluents. Generally, such carriers should be nontoxic to recipients at the dosages and concentrations employed. Ordinarily, the preparation of such compositions entails combining the therapeutic agent with buffers, antioxidants such as ascorbic acid, low molecular weight (less than about 10 residues) polypeptides, proteins, amino acids, carbohydrates including glucose, sucrose or dextrans, chelating agents such as EDTA, glutathione and other stabilizers and excipients. Neutral buffered saline or saline mixed with nonspecific serum albumin are exemplary appropriate diluents.

In addition, the pharmaceutical compositions of the present invention may be prepared for administration by a variety of different routes. In addition, pharmaceutical compositions of the present invention may be placed within containers, along with packaging material which provides instructions regarding the use of such pharmaceutical compositions. Generally, such instructions will include a tangible expression describing the reagent concentration, as well as within certain embodiments, relative amounts of excipient ingredients or diluents (*e.g.*, water, saline or PBS) which may be necessary to reconstitute the pharmaceutical composition.

Methods of Treating or Preventing Werner's Syndrome

The present invention also provides methods for treating or preventing Werner's Syndrome (or related diseases), comprising the step of administering to a patient a vector (*e.g.*, expression vector, viral vector, or viral particle containing a vector) or nucleic acid molecules alone, as described above, thereby reducing the risk of developing Werner's Syndrome. The present invention also provides a method for using the present invention to delay onset of Werner's Syndrome, lessen symptoms, or hasten or delay

progression of the disease. Such therapeutics may be tested in a transgenic animal model that expresses mutant protein, wild-type and mutant protein, or in an *in vitro* assay system (e.g., a helicase assay such as that described by Bjornson et al., *Biochem. 3307:14306-14316*, 1994).

5 As noted above, the present invention provides methods for treating or preventing Werner's Syndrome through the administration to a patient of a therapeutically effective amount of an antagonist or pharmaceutical composition as described herein. Such patients may be identified through clinical diagnosis based on the classical symptoms of Werner's Syndrome.

10 As will be evident to one of skill in the art, the amount and frequency of administration will depend, of course, on such factors as the nature and severity of the indication being treated, the desired response, the condition of the patient, and so forth. Typically, the compositions may be administered by a variety of techniques, as noted above.

15 Within other embodiments of the invention, the vectors which contain or express the nucleic acid molecules which encode the WRN proteins described above, or even the nucleic acid molecules themselves may be administered by a variety of alternative techniques, including for example administration of asialoosomucoid (ASOR) conjugated with poly-L-lysine DNA complexes (Cristano et al., *PNAS* 92:122-
20 92126, 1993), DNA linked to killed adenovirus (Curiel et al., *Hum. Gene Ther.* 3(2):147-154, 1992), cytofectin-mediated introduction (DMRIE-DOPE, Vical, Calif.), direct DNA injection (Acsadi et al., *Nature* 352:815-818, 1991); DNA ligand (Wu et al., *J. of Biol. Chem.* 264:16985-16987, 1989); lipofection (Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417, 1989); liposomes (Pickering et al., *Circ.* 89(1):13-21, 1994;
25 and Wang et al., *PNAS* 84:7851-7855, 1987); microprojectile bombardment (Williams et al., *PNAS* 88:2726-2730, 1991); and direct delivery of nucleic acids which encode the WRN protein itself either alone (Vile and Hart, *Cancer Res.* 53: 3860-3864, 1993) or

The following examples are offered by way of illustration, and not by way of limitation.

EXAMPLES

EXAMPLE 1

CLONING OF THE WRN GENE FROM CHROMOSOME 8

5

The WS locus (*WRN*) was initially localized to 8p12 by conventional mapping methods (Goto et al., *Nature* 355:735-738, 1992) and the genetic position refined using both meiotic and homozygosity mapping (Schellenberg et al., 1992; Nakura, et al., *Genomics* 23:600-608, 1994; Thomas, *Genomics* 16:685-690, 1993).

10 The latter approach is possible since many WS subjects are the offspring of consanguineous marriages (Table 1). Initial mapping work (Nakura, et al., *Genomics* 23:600-608, 1994; Oshima et al., *Genomics* 23:100-113, 1994) placed the *WRN* locus in an 8.3 cM interval flanked by *D8S137* and *D8S87* (Fig. 1). *D8S339*, a marker within this interval, was the closest locus tested ($q = 0.001$, $Z_{\max} = 15.93$). Multipoint analysis
15 placed *WRN* within 0.6 cM of *D8S339*, although the region between *D8S87* and *FGFR* could not be excluded. Subsequently, the short tandem repeat polymorphism (STRP) markers at glutathione reductase (*GSR*) and *D8S339* were found to be in linkage disequilibrium with WS in Japanese WS subjects (Yu, *American Journal of Human Genetics* 55:356-364, 1994).

20 To clone the *WRN* gene, a yeast artificial chromosome (YAC) P1, and cosmid contig was generated starting at the *GSR/D8S339* region and extended by walking methods to cover approximately 3 Mb. An additional 16 STRP markers in the YAC contig (Fig. 1B) were identified to define recombinants and to delineate the boundaries of the linkage disequilibrium region. For marker ordering and gene
25 identification, cosmids and P1 clones were also isolated and used to construct a small-clone partial contig of the region (Fig. 1E). The *WRN* region was defined by obligate recombinants at C41C3S3 excluding the region telomeric to this marker, and at C896P9

30 region

Genes in the WRN region were identified by exon trapping using vector pSL3 (Buckler et al., *Proc. Natl. Acad. Sci. USA* 88:4005-4009, 1991; Church et al., *Nat. Genet.* 6:98-105, 1994), hybridization of cDNA libraries to immobilized YACs (Parimoo et al., *Proc. Natl. Acad. Sci. USA* 87:3166-3169, 1991), and comparison of the genomic sequence to DNA sequence databases using BLAST (Altschul et al. *J. Mol. Biol.* 215:403-410, 1990) and the exon-finding program GRAIL (Uberbacher and Mural, *Proc. Natl. Acad. Sci. USA* 88:1261, 1991). The genomic sequence was determined for the region defined by P1 clones 2233, 2253, 3833, 2236, 2237, 2932, 6738 and 2934 and cosmid clone 176 C6. Each method identifies short segments of expressed sequences, which were then used to screen an arrayed fibroblast cDNA library to identify longer cDNA clones. This library was selected because WS fibroblasts have a premature senescence phenotype *in vitro*, indicating that the WRN gene is probably expressed in this cell type. Genes identified by this process were screened for WRN mutations using reverse transcriptase-polymerase chain reaction (RT-PCR). Seven subjects were initially screened for mutations: 5 WRN subjects (2 Caucasians and 3 Japanese) and 2 control subjects (1 Caucasian and 1 Japanese). Prior to identification of the WRN gene, the following genes from the region were screened for mutations: GSR, PP2AB, TFIIEB, and genes corresponding to other expressed sequence tagged sites (ESTs).

The candidate WRN locus gene was initially detected by using the genomic sequence of P1 clone 2934 to search the EST database. A single 245 bp EST, R58879, was detected which is homologous to 3 segments of the genomic sequence separated by presumed intronic sequence. Sequence from R58879 was used to identify longer cDNA clones from a normal fibroblast cDNA library. An initial 2.1 kb cDNA clone containing EST R58879, which corresponds to the 3' end of the gene, was obtained by screening an array of clones by PCR, using the primers A and B (see below). Primers A and B are derived from R58879 sequence and yield a 145 bp

located in p2934 and to sequences contained in the initial 2.1 kb clone. Six additional

clones were identified. An additional 8 clones were obtained by plaque hybridization. The longest clone is 4.0 kb in length. Additional sequence was obtained by the RAGE method using primer 5EA to prime first strand cDNA synthesis. A 2.5 kb product was obtained that contained an additional 1.4 kb of sequence.

5 Evidence that R58879 is expressed was obtained by Northern blot analysis, in which 6.5 kb and 8 kb transcripts were detected in a variety of tissues, including heart, placenta, muscle, and pancreas. Also, transcripts were detected by RT-PCR products from fibroblast and lymphoblastoid cell line RNA.

10 EXAMPLE 2

CLONING OF THE WRN GENE FROM SUBJECTS

The WRN gene may be isolated from patients and mutations or polymorphisms determined by sequence analysis. Peripheral blood cells are obtained
15 by venipuncture and hypotonic lysis of erythrocytes. DNA or RNA is isolated from these cells and the WRN gene isolated by amplification. The gene sequence may be obtained by amplification of the exons from genomic DNA or by RT-PCR, followed by determination of the DNA sequence. Primers suitable for determining the DNA sequence and for performing RT-PCR are listed below (Primers A-R are SEQ ID Nos.
20 1-18 respectively, and primers 5EA-5EG are SEQ ID Nos. 19-25 respectively). Two cDNAs were identified and are shown in Figures 2 and 3. There is some uncertainty regarding the identity of a few bases in the 5' untranslated region in Figure 2

Two RT-PCR reactions are used to obtain the gene from different tissues. First strand cDNA synthesis is carried out according to standard procedures
25 (e.g., with a Stratascript Kit from Stratagene). The cDNA is subjected to a pair of nested PCR amplifications, the first with primers I and J (SEQ ID Nos. 9 and 10), followed by primers K and L (SEQ ID Nos. 11 and 12), and the second with primers

30 primer H and G, and then sequencing. Identifying differences in

the gene sequence or splicing pattern. Primers A-H (SEQ ID Nos. 1-8) and K-R (SEQ

ID Nos. 11-18) are used for sequencing the first RT-PCR fragment. Primers B, 5EA, 5EB, 5EC, 5EE, 5EF and 5EG (SEQ ID Nos. 2, 19, 20, 21, 23, 24, and 25, respectively) are used for sequencing the second RT-PCR fragment. Sequencing is done on an ABI373A using Applied Biosystems Division of Perkin-Elmer FS sequencing kits according to the instructions of the manufacturer.

A	5'-CTGGCAAGGATCAAACAGAGAG
B	5'-CTTTATGAAGCCAATTTCTACCC
C	5'-TGGCAAATTGGTAGAAGCTAGG
10 D	5'-AAATAACTATGCTTTCTTACATTTAC
F	5'-CTCCCGTCAACTCAGATATGAG
F	5'-CTGTTTGTAATGTAAGAAAGCATAG
G	5'-GAGCTATGATGACACCACTGC
H	5'-ACTGAGCAACAGAGTGAGACC
15 I	5'-GGATCTGGTCTCACTCTGTTGC
J	5'-TTGCCTAGTGCAATTGGTCTCC
K	5'-AGTGCAGTGGTGTATCATAGC
L	5'-CCTATTTAATGGCACCCAAAATGC
M	5'-CAGTCTATGGCCATCACATACTC
20 N	5'-ACCGCTTGGGATAAGTGCATGC
O	5'-GAGAAGAAGTCTAACTTGGAGAAG
P	5'-TTCTGGTGACTGTACCATGATAC
Q	5'-CCAAAGGAAGTGATACCAGCAAG
R	5'-ACAGCAAGAAACATAATTGTTCTGG
25 5EA	5'-GAACTTTGAAGTCCATCACGACC
5EB	5'-GCATTAATAAAGCTGACATTGCGC
5EC	5'-CATTACGGTGCTCCTAAGGACATG
5ED	5'-GATGGATTTGAAGATGGAGTAGAAG
5EE	5'-TGAAAGACAATATGGAAAGAGCTTG
30 5EF	5'-GTAGAACCAACTCATTCTAAATGCT
5EG	5'-AATTGCGTGTCTCCTTGCGCA

The exons of the 3'-end of the WRN gene can be amplified from DNA samples using the primers listed below (Primers E1A-E13B are SEQ ID Nos. 26-57, respectively). The DNA sequence is determined using the same primers and an

E1A 5'-CTTATAGTCAAGGATGAGTTC

```

E1B 5'-CATGAAACTTGCTTCTAGGACAC
E2A 5'-CCCAGGAGTTGAGACCATCC
E2B 5'-TTACAATCGGCCACATTCATCAC
E2C 5'-TGTAATCCCAACACTTTGGGAGG
5 E2D 5'-AGTGAAGAATTCATAGTGGATGG
E3A 5'-TAGCTTTATGAAGCCAATTTCTACC
E3B 5'-AATCCAAAGAATCAATAGACAAGTC
E3C 5'-GCTTGAAGGATGAGGCTCTGAG
E3D 5'-TGTCAGAAATGAGCACCATGGG
10 E4A 5'-CTTGTGAGAGGCCTA1AAACTCG
E4B 5'-GGTAAACAGTGTAGGAGTCTGC
E5A 5'-GCCATTTTCTCTTAATTGGAAAGG
E5B 5'-ATCTTATTCATCTTTCTGAGAATGG
E6A 5'-TGAAATAGCCCAACATCTGACAG
15 E6B 5'-GATTAATTTGACAGCTTGATTAGGC
E7A 5'-TGAAATATAAACTCAGACTCTTAGC
E7B 5'-GTAATGATTTGGAAAGACATTCTC
E8A 5'-GATGTGACAGTGAAGCTATGG
E8B 5'-GGAAAAATGIGGTATCTGAAGCTC
20 E9A 5'-AAGTGAGCAAATGTTGCTTCTGG
E9B 5'-TCATTAGGAAGCTGAACATCAGC
E10A 5'-GTTGGAGGAAATTGATCCCMAGTC
E10B 5'-TGTTGCTTATGGGTTTAACTTGTG
E11A 5'-TAAAGGATTAATGCTGTTAACAGTG
25 E11B 5'-TCACACTGAGCATTTACTACCTG
E12A 5'-GTAA1CATA1CAGAATTCATAACAG
E12B 5'-CTTTGGCAACCTTCCACCTTCC
E12C 5'-GCAAMGGAAATGTAGCACATAGAG
E12D 5'-AGGCTATAGGCATTTGAAAGAGG
30 E13A 5'-GTAGGCTCCAGAAGACCCAG
E13B 5'-GAAAGGATGGGTGTGTATTCAGG

```

EXAMPLE 3

35

IDENTIFICATION OF MUTANT ALLELES

The cDNA sequence (Figure 2) was aligned to the genomic sequence to

identify mutations.

In four of the five patients, single base pair changes lead to splicing defects or stop codons in the open reading frame of the gene. In the fifth patient, a single base pair

change results in a cysteine to arginine transition, which may disrupt gene function. Each of the exons was also sequenced in 96 unaffected control individuals (48 Caucasians and 48 Japanese), and none of the mutations were found in any of the control individuals.

- 5 The first mutation is a mutation at a splice acceptor site. In the sequence below, the GGTAGAAA sequence begins at nucleotide 2030 (Figure 2). The g to c change results in a deletion of 95 bp.

- Preparation of DNA for RT-PCR mutational analysis revealed that for one subject, the amplification product was shorter than observed in products from other
10 WS and control subjects. DNA sequence analysis of the RT-PCR product revealed that 95 bp were missing compared to other samples. The missing sequence corresponds to a single exon. This exon and flanking genomic segments were sequenced from the WS subject and controls and a single base change (G→C) at the splice donor site was detected. The subject was the offspring of a first cousin marriage and was, as expected,
15 homozygous for this mutation. The same mutation was found in a total of 18 out of 30 Japanese WS subjects and, thus, is the most common Japanese WS mutation. Deletion of this exon results in a change in the predicted open-reading frame and a premature stop codon. This mutation was not observed in 46 Japanese and 46 Caucasian controls. Among mutation carriers, 12/16 had the 141 bp allele at the GSR2-STRP.

20

wild type: ttttaatatagGGTAGAAA	(SEQ ID No. 58)
Werners: ttttaatacGGTAGAAA	(SEQ ID No. 59)

- The second mutation changes a C to T at nucleotide 2384 (Figure 2)
25 changing a glutamine to a stop codon, which results in a predicted truncated protein. This mutation was observed in a single subject. Primers E11A and E11B flank this sequence and amplify a 360 bp fragment.

primer	5'	AG	3'	stop
		<u>T</u>		
		ter		

The third mutation changes a C to T at nucleotide 2804 (Figure 2), which alters an arginine codon to a stop codon resulting in a predicted truncated protein. Four Japanese WS subjects and 1 Caucasian W5 subject had this mutation. Primers E8A and E8B flank this sequence and amplify a 267 bp product.

arg
wild type: TTGGAGCGAGCA (SEQ ID No. 62)
Werners: TTGGAGTGAGCA (SEQ ID No. 63)
ter

The fourth mutation is a 4 bp deletion across a splice junction. The exon sequence shown below begins at nucleotide 2579 (Figure 2). This mutation was identified in a Syrian W5 kindred. Primers E4A and E4B flank this mutation and amplify a 267 bp fragment.

wild type: ctgt**ag**ACAGACACCTC (SEQ ID No. 68)
Werners: ctgt----AGACACCTC (SEQ ID No. 69)

The fifth mutation is a missense mutation. A T is altered to a G at nucleotide 2113 (Figure 2), changing the wild-type phe codon to a leu codon. This change is a polymorphism with each allele present at a frequency of approximately 0.5. It does not appear to correlate with WS.

phe
wild type: AAGAAGTTTCTTCTG (SEQ ID No. 64)
Werners: AAGAAGTTGCTTCTG (SEQ ID No. 65)
leu

The sixth mutation is a missense mutation changing a T to a C at

wild type: CTTCATGTGAT (SEQ ID No. 66)
Werners: CTTCACTGAT (SEQ ID No. 67)

45

arg

These point mutations may also be identified by PCR using primers that contain as the 3'-most base either the wild type or the mutant nucleotide. Two separate
5 reactions are performed using one of these primers and a common second primer. Amplification is detectable in the reaction containing a matched primer.

EXAMPLE 4

10 CHARACTERIZATION OF THE WRN GENE AND GENE PRODUCT

The 2 kb WRN cDNA hybridizes to a 6.5 kb RNA and a less abundant 8 kb RNA on a Northern blot, suggesting that a full length coding region is about 5.2 kb long. An overlapping cDNA clone has been isolated that extends the sequence by 2 kb.
15 The insert from this clone is used to probe cDNA libraries to identify other clones that contain the 5' end of the cDNA or full length sequence. Alternate splicing events are detected by sequencing the full cDNA sequence from a number of different tissues, including fully differentiated cells and stem cells, and the full range of gene transcripts identified by sequence comparison. Additional exons are identified as above by further
20 genomic sequencing and GRAIL analysis.

The predicted amino acid sequence is shown in Figures 2B and 3. Figure 2 shows cDNA and predicted amino acid sequences of the WRN gene. Figure 3 presents cDNA and predicted amino acid sequences of a less abundant transcript of the WRN gene. The longest open reading frame is shown from the first methionine in that
25 frame. The predicted WRN protein consists of 1,432 amino acids divided into three regions: an N-terminal region, a central region containing 7 motifs (I, Ia, II, III, IV, V and VI) characteristic of the DNA and RNA superfamily of helicases (Gorbalenya et al. *Nucleic Acid Res.* 17: 1713-1989, 1989; Gorbalenya et al. *EMBO J.* 8: 1473-1481, 1989).

Now, having described the invention, it will be clear to those skilled in the art how to make and use the same. Because many helicases function as part of a multiprotein complex, the N-terminal

and/or the C-terminal domain may contain interaction sites for these other proteins, while the central helicase domain functions in the actual enzymatic unwinding of DNA or RNA duplexes.

The N-terminal region, encompassing approximately codons 1 to 539, is
5 acidic; there are 109 aspartate or glutamate residues, including a stretch of 14 acidic residues in a 19 amino acid sequence (codons 507-526). Stretches of acidic residues are found in the Xeroderma pigmentosum (XP) complementation group B helicase, the Bloom's syndrome helicase, and the X-chromosome-linked α -thalassemia mental retardation syndrome helicase. In the WRN gene, this region also contains a tandem
10 duplication of 27 amino acids in which each copy is encoded by a single exon. Because this duplication is exact at the nucleotide level, and because flanking intronic sequences for the two exons that encode the duplication are also highly similar, this duplication is presumed to be the result of a relatively recent event. The duplicated regions are also highly acidic with 8 glutamate or aspartate residues out of 27 amino acids and only 2
15 basic amino acids (one histidine and one lysine residue).

The central region of the WRN gene, spanning approximately codons 540-963, is highly homologous to other helicases from a wide range of organisms including the ReqQ gene from *E. coli*, the SGS1 gene from *S. cerevisiae*, a predicted helicase (F18C5C) from *C. elegans*, and several human helicases. Thus, by sequence
20 similarity, the WRN gene is a member of a superfamily of DExH-box DNA and RNA helicases. The principle conserved sequences consist of 7 motifs found in other helicases. These motifs include a predicted nucleotide binding site (motif I) and a Mg²⁺ binding site (sequence DEAH, motif II). Some or all of the 7 motifs are presumed to form the enzymatic active site for DNA/RNA unwinding. The presence of the DEAH
25 sequence and an ATP-binding motif further suggests that the WRN gene product is a functional helicase.

The C-terminal end of the WRN gene, from codons 964 to 1432, has

EXAMPLE 5

IDENTIFYING AND DETECTING MUTATIONS IN THE WRN GENE

5

Mutations or polymorphisms of WRN may be identified by various methods, including sequence analysis. Although any cell (other than erythrocytes) may be used to isolate nucleic acids, peripheral blood mononuclear cells (PBMC) are preferred. Peripheral blood mononuclear cells are obtained by venipuncture and subsequent hypotonic lysis of erythrocytes. RNA is isolated and first strand cDNA synthesis is performed using a Strata-script RT-PCR kit according to the manufacturers instructions (Stratagene, La Jolla, part numbers 200347 and 200420). Three RT-PCR fragments are amplified using an LA PCR Kit Ver. 2 using buffer containing 1.5 mM Mg+2 (TaKaRa Shuzo Co., Ltd., Japan, part number RR013A). Nested PCR is performed. In this reaction, a second PCR is performed using a pair of primers within the sequence amplified by the first PCR reaction. The cycling conditions for each amplification are: 10 min at 95°C, 35 cycles of 1 min at 60°C, 1 min at 72°C, and 1 min at 95°C, followed by 7 min at 72°C in a Perkin-Elmer 9600 PCR machine. The amplified fragments are purified using 96-well plate spin columns (Wang et al., *Anal. Biochem.* 226:85-90, 1995). DNA sequence is determined using an FS Dye-Terminator sequencing kit (Applied Biosystems Division of Perkin Elmer) and the specific primers described below. An automated Applied Biosystems ABI373A DNA Sequencer is used to determine the sequence. The amplified fragments and the appropriate primers are listed in Table 1, and the primer sequences are listed in Table 2.

25

The DNA sequences are aligned with the known sequence (Figure 2A) using the program Sequencher (Gene Codes, Michigan) to identify any discrepancies between patient samples and the reference sequence.

Table 1 PCR and sequence primers

Fragment	Primers Nested on cDNA		Coordinates	Sequence primers
	1st PCR	2nd PCR		
I	5EC, J	5EN, L	2947-5065	5EN, L, M, N, O, P, Q, R
II	5ED, P	5EE, B	1379-3391	5EE, 5EJ, 5EK, 5EL, 5EM, 5EB, 5EA, 5EN, B
III	5ES, 5EK	5ET, 5EH	75-1516	5ET, 5EX, 5E1, 5EP, 5EO, 5ED, 5EH

Table 2 Primer sequences

5	B	5'-CTTTATGAAGCCAATTTCTACCC	(SEQ ID No. 2)
	J	5'-TTGCCTAGTGCAATTGGTCTCC	(SEQ ID No. 10)
	L	5'-CCTATTTAATGGCACCACAAATGC	(SEQ ID No. 12)
	M	5'-CAGTCTATGGCCATCACATACTC	(SEQ ID No. 13)
10	N	5'-ACCGCTTGGGATAAGTGCATGC	(SEQ ID No. 14)
	O	5'-GAGAAGAAGTCTAACTTGGAGAAG	(SEQ ID No. 15)
	P	5'-TTCTGGTGACTGTACCATGATAC	(SEQ ID No. 16)
	Q	5'-CCAAAGGAAGTGATACCAGCAAG	(SEQ ID No. 17)
	R	5'-ACAGCAAGAAACATAATTGTTCTGG	(SEQ ID No. 18)
15	5EA	5'-GAACTTTGAAGTCCATCACGACC	(SEQ ID No. 19)
	5EB	5'-GCATTAATAAAGCTGACATTCGCC	(SEQ ID No. 20)
	5EC	5'-CATTACGGTGCTCCTAAGGACATG	(SEQ ID No. 21)
	5ED	5'-GATGGATTTGAAGATGGAGTAGAAG	(SEQ ID No. 22)
	5EE	5'-TGAAAGAGAATATGGAAAGAGCTTG	(SEQ ID No. 23)
20	5EH	5'-CATTGGGAGATAAATGCTCAGTAGA	(SEQ ID No. 80)
	5EJ	5'-AGATGTACTTTGGCCATTCCAG	(SEQ ID No. 81)
	5EK	5'-GCCATGACAGCAACATTATCTC	(SEQ ID No. 82)
	5EL	5'-CTTACTGCTACTGCAAGTTCTTC	(SEQ ID No. 83)
	5EM	5'-TCGATCAAAACCAAGTACAGGTG	(SEQ ID No. 84)
25	5EN	5'-GCAGATGTAGGAGACAAATCATC	(SEQ ID No. 85)
	5EO	5'-TCATCCAAAATCTCTAAATTTCTGG	(SEQ ID No. 86)
	5EP	5'-CTGAGGACCAGAACTGTATGC	(SEQ ID No. 87)
	5ES	5'-GCTGATTTGGTGTCTAGCCTGG	(SEQ ID No. 88)
	5ET	5'-TGCCTGGGTTGCAGGCCTGC	(SEQ ID No. 89)
30	5EX	5'-TTGGAAACAACCTGCACAGCAGC	(SEQ ID No. 90)

EXAMPLE 6

ISOLATION OF GENOMIC DNA CONTAINING WERNER'S SYNDROME GENE

5

To facilitate mutational analysis of the *WRN* gene, the intron-exon structure is determined. The *WRN* gene is located in the genomic sequence of P1 clone 2934. However, this clone only contains the 3' end of the gene (exons 21 to 35). Genomic clones containing the 5' end are obtained from a chromosome 8-specific cosmid library LA08NC01 (Wood et al. *Cytogenet. Cell Genet.* 59: 243, 1992) by
 10 screening for clones adjacent to P1 clone 2934. Briefly, this library is arrayed for PCR screening as described in Amemiya et al. (*Nucl. Acids Res.* 20: 2559, 1992). *WRN* containing cosmids are identified using primer sets 5E6/5EY, 5ED/5E12, and CD-A/CD-B (Table 3), which are derived from the *WRN* cDNA sequence (Figure 1;
 15 GenBank Accession No. L76937). Four walking steps yielded cosmids 193B5, 114D2, 78D8 and 194C3, which contained the remaining exons. Primers derived from the *WRN* cDNA were used for the initial sequence analysis of the cosmid clones. The resulting sequence (Figure 5) is compared to the cDNA sequence to identify intron-exon boundaries. Sequencing primers are then designed from the intron sequences to obtain
 20 sequence in the reverse direction and to obtain the second boundary defining the intron-exon junction. This strategy is used to define the exons not present in P1 clone 2934.

Table 3. Primer sequence and PCR conditions for *WRN* analysis

Region	Primer Sequence	Product Size (bp)	Mg ⁺² (mM)	pH
N-domain	5E6 5'-GATATTGTTTGTATTTACCCATGAAGAC (SEQ ID No. 164) 5EY 5'-TCCGCTGCTGTGCAGTTCTTTCC (SEQ ID No. 165)	106	1.5	8.3

The annealing temperature was 60° C for all primer sets.

Table 4 presents a summary of the structure of the genomic WRN gene.

- 5 The first column identifies the exon, the second column indicates the base numbers of the cDNA that are derived from the exon, the third column denotes the size of the exon in bp, the fourth column shows the sequence of the boundaries with intron sequences in lower case letters and exon sequences in upper case letters, the fifth column shows notable features of the exons.

10

Table 4. Intron-Exon Structure of the *WRN* Gene

Exon	cDNA Location	Exon Size (bp)	Intron-Exon Boundary Sequences	Exon Features
1	1-155	>155	...TTCTCGGGgtaaagtgtc (SEQ ID No. 169)	5' UTR
2	156-327	172	tacctctcagTTTTCTT...AAAGAAAGgtatgttgtt (SEQ ID No. 170)	5' UTR, ATG codon
3	328-440	113	taaactcaagGCATGTGT...GATATTAGgtaagtgatt (SEQ ID No. 171)	
4	441-586	146	ctcacttttagCATGAGTC...CATGTCAGgttggtatct (SEQ ID No. 172)	
5	587-735	149	aatgttacagTTTTCCC...ATAAAAAGgtaaaagcaa (SEQ ID No. 173)	
6	736-885	150	tcatttctagCTGAATG...ATGCTTAgttacgttctt (SEQ ID No. 174)	
7	886-955	70	ttttttataagGCTGGTTT...AAATAAGgtatgtttaa (SEQ ID No. 175)	
8	956-1070	115	ttccccctagAGGAAGAA...CCACGGAGgttaaataatt (SEQ ID No. 176)	
9	1071-1500	430	tttttttagGGTTTCTA...CTACTGAGgtactaaaat (SEQ ID No. 177)	
10	1501-	81	tttttaaaacATTATC...TGCTTAAGggtatgtttt (SEQ ID No. 178)	3' UTR, stop codon
11	1503-1807	305	ttacttttagGCTTAAAG...GATAAGgtttttttttt (SEQ ID No. 180)	

Exon	cDNA Location	Exon Size (bp)	Intron-Exon Boundary Sequences	Exon Features
13	1808-1883	76	ttattttcagACTTTTGG...TTTAAACCgtgagctataa (SEQ ID No. 181)	
14	1884-1951	68	caccttcaagAGTTCACT...GGCAACTGgtaagtgtga (SEQ ID No. 182)	helicase motif I (5' end)
15	1952-2060	109	tcatttcaagGATATGGA...CAGCTTAAgtaagtcacg (SEQ ID No. 183)	helicase motif I (3' end) and Ia
16	2061-2129	69	cttccttatagAATGTCCA...ATTAAATTgtgagtaatt (SEQ ID No. 184)	
17	2130-2212	83	gtttttacagAGGTAAAT...TGATATTGgtaagtgtga (SEQ ID No. 185)	
18	2213-2319	107	ttttttacagGTATCAGG...TGCCAATGgtaagctttg (SEQ ID No. 186)	helicase motif II
19	2320-2504	185	catcattcagGTTCCAAT...AAAACAAGgtaaggattt (SEQ ID No. 187)	helicase motif III
20	2505-2679	175	ttttcttttagTTCCCACT...AAATTCAGgtatgaggat (SEQ ID No. 188)	helicase motif IV
21	2680-2861	182	ttgtttctcagTGTGTCT...TTAAATAGgtaaaaaaa (SEQ ID No. 189)	helicase motifs V and VI
22	2862-2963	102	taatgcacagGCACCTTC...AGGAGACAgtatgtatta (SEQ ID No. 190)	
23	2964-3056	93	tcttgggtagAATCATCT...AGGTCCAGgtaaaagattt (SEQ ID No. 191)	
24	3057-3198	142	ttttatttagATTGGATC...GAGGATCTgtaagtatat (SEQ ID No. 192)	
25	3199-3369	171	ctaatttcagAATCTCA...CGAAAAAGgtaaacagtg (SEQ ID No. 193)	
26	3370-3464	95	cttttaatagGGTAGAAA...CTGCCTAGgtttaaatttt (SEQ ID No. 194)	
27	3465-3540	76	tttttttttagTTGAAAA...AGAAGAAGgtttgtttta (SEQ ID No. 195)	
28	3541-3614	74	ttaaatgcagTCTAAGTT...AAAAAAGgtacagagtt (SEQ ID No. 196)	
29	3615-3690	76	aatatttttagTATCATGG...AGACTCAGgtaaaacatt (SEQ ID No. 197)	

FIG. 1 shows the sequence of the cDNA for the gene identified as 197.

Exon	cDNA Location	Exon Size (bp)	Intron-Exon Boundary Sequences	Exon Features
32	3919-4050	132	aattctgtagACAGACCT...TGCCCTTGgtaagtgtga (SEQ ID No. 200)	
33	4051-4213	163	ctttctctagAAGAGCAT...CAACTCAGgtgagaggca (SEQ ID No. 201)	
34	4214-4422	209	tcgtttacagATATGAGT...ATACTGAGgtattaatta (SEQ ID No. 202)	
35	4423-5190	768	tttctacagACTTCATC... (SEQ ID No. 203)	TAA codon, 3'UTR

Note. Exons are in uppercase and intron sequences are in lowercase letters.

As shown above, WRN contains a total of 35 exons ranging in size from 68 bp (exon 14) to 768 bp (exon 35). The coding region begins in the second exon (Table 2). As noted previously, there is a duplicated region in the WRN cDNA sequence which is 27 amino acids in length. This duplication is exactly conserved at the nucleotide level in cDNA. At the genomic level, the duplicated sequences were present as 2 exons (exons 10 and 11), each exon containing only the duplicated nucleotides. The intronic sequences adjacent to these 2 exons are also highly conserved, suggesting that the a relatively recent duplication event is responsible for these repeated exons. In addition, because the surrounding intronic sequences were conserved, it was not possible to design primers which could specifically amplify exons 10 and 11.

The helicase region of the WRN gene is contained in exons 14-21. Helicase motif 1 is split between exons 14 and 15 while the remaining motifs are each in an individual exon (Table 4). This region, from codon 569 to 859, has sequence similarity to the 7 signature helicase motifs. In addition, though the sequences between the motifs are not conserved, the spacing is very similar in genes from a wide range of species. For example, the helicase domains in the *E. coli RecQ* gene are found in a

EXAMPLE 7

IDENTIFICATION OF MUTATIONS

Initially, 4 different mutations in the C-terminal domain of *WRN* were identified. These mutations accounted for more than 80% of the Japanese WS patients examined. All 4 mutations are in the C-terminal domain region of *WRN* and the resulting predicted protein contained an intact helicase domain. Additional WS subjects are screened to identify further mutations. Genomic structure information is used to design PCR-primers for amplifying each exon, which is then subjected to DNA sequence analysis. Five additional *WRN* mutations are described; 2 are located in the consensus helicase motifs and another 2 are predicted to produce truncated proteins without the helicase domains. These mutations suggest that in at least some WS subjects, the enzymatic helicase activity is destroyed and support that complete loss-of-function of *WRN* gene product causes Werner's syndrome.

Although any cell may be used to isolate DNA, PBMC are preferred. As above, PBMC are obtained by venipuncture and subsequent hypotonic lysis of erythrocytes. PBMC are lysed by the addition of detergent, such as 0.5% NP-40, 0.5% Triton-X100, or 0.5% SDS. If a non-ionic detergent is used, no further purification of DNA is necessary, but proteinase K treatment, and subsequent heat killing of the enzyme (95°C for 10 minutes) is required. Genomic DNA is amplified according to the PCR conditions recited above using the primers listed in Table 5. Exons 9 and 10 are contained in a region of DNA that is duplicated. The primer pair for exon 9 and 10 anneals to sequences outside the duplication. Amplified product is analyzed by DNA sequence determination, hybridization with allele-specific probe, or other mutation detection method. When DNA sequences are determined, the sequence of the amplified exon is aligned with the known sequence (Figure 2A) and any discrepancies between patient samples and the reference sequence are identified.

Table 5

PCR Fragment	Primer Sequence	Product Size (bp)	Mg ²⁺ (mM)	pH
exon 1	A 5'-AGGGCCTCCACGCATGACGC (SEQ ID No. 92) B 5'-AGTCTGTTTTCCAGAATCTCCC (SEQ ID No. 93)	583	1.5	8.3
exon 2	A 5'-CCTATGCTTGGACCIAGGTGTC (SEQ ID No. 94) B 5'-GAAGTTTACAAGTAACAACAGTACIC (SEQ ID No. 95)	339	1.5	8.3
exon 3	A 5'-ACTATAAATTGAATGCTTCAGTGAAC (SEQ ID No. 96) B 5'-GAACACACCTCACCTGTAAAACTC (SEQ ID No. 97)	316	1.5	8.3
exon 4	E 5'-GGTAAACCACCATACCTGGCC (SEQ ID No. 98) F 5'-GTACATATCCTGGTCATTTAGCC (SEQ ID No. 99)	691	1.5	8.3
exon 5	B 5'-ATTCAGATAGAAAGTACATCTGTG (SEQ ID No. 100) E 5'-GTTAAGAAATACTCAAGGTCAATGTG (SEQ ID No. 101)	369	1.5	8.3
exon 6	A 5'-GGTTGTATTTTGGTATAACATTTCC (SEQ ID No. 102) B 5'-ATATTTTGGTAGAGTTTCTGCCAC (SEQ ID No. 103)	374	1.5	8.3
exon 7	A 5'-CTCTTCGATTTTTCTGAAGATGGG (SEQ ID No. 104) B 5'-CCCTAATAGTCAGGAGTGTCAAG (SEQ ID No. 105)	291	1.5	8.3
exon 8	A 5'-GGAAAGAAAATGAAAATTTGATCCC (SEQ ID No. 106) B 5'-CAGCCTTAATGAATAGTATTCTTCAC (SEQ ID No. 107)	316	4.0	8.3
exon 9	C 5'-ATIGATCTTTAAGTGAAGGTGAGC (SEQ ID No. 108) D 5'-CAACCAAAATATCAAAAATTGGATC (SEQ ID No. 109)	668	1.5	8.3

PCR Fragment	Primer Sequence	Product Size (bp)	Mg ²⁺ (mM)	pH
	(SEQ ID No. 111)			
exon 13	A 5'-TAACCCATGGTAGCTGCACTG (SEQ ID No. 112) B 5'-CTGTTGCTGTTAAGCAGACAGG (SEQ ID No. 113)	285	1.5	8.3
exon 14	C 5'-TTGAATGGGACATTGSTCAAATGG (SEQ ID No. 114) F 5'-GTAGTTGCAITTTGTATTTGAGAGT (SEQ ID No. 115)	348	1.5	8.3
exon 15	G 5'-GTAAAAAGAAATGAAAGCATCAAAGG (SEQ ID No. 116) D 5'-TCACCCACAGAAGAAAAAGAGG (SEQ ID No. 117)	246	4.0	8.3
exon 16	A 5'-CAAAAAAGAAAAATTGCAAGAACAGG (SEQ ID No. 118) B 5'-CAGCAACATGTAATTCACCCACG (SEQ ID No. 119)	282	4.0	8.3
exon 17	S 5'-GAAGAGACTGGAATTGGGTTTGG (SEQ ID No. 120) S 5'-ATAGAGTATCATGGGATAAGATAGG (SEQ ID No. 121)	532	1.5	8.3
exon 18	A 5'-TTCTCCTTTGGAGATGTAGATGAG (SEQ ID No. 122) B 5'-TCTTCAGCTTCTTTACCAGTCCCA (SEQ ID No. 123)	273	4.0	10
exon 19	A 5'-CATGGTGTTTGACAACAGGATGG (SEQ ID No. 124) B 5'-GTTAAATATGCATTAGAAGGAAATCC (SEQ ID No. 125)	396	4.0	9.0
exon 20	A 5'-ATAAAACCAAACGGGTCTGAAGG (SEQ ID No. 126) B 5'-AAAAGAAGTATTCAATAAGATCTGG (SEQ ID No. 127)	342	4.0	8.3
exon 21	A 5'-AATTCACCTTTGTGCCAGGGACT (SEQ ID No. 128) B 5'-ACTTGGGATAGTCAAAATAGCTT (SEQ ID No. 129)	397	1.5	9.0

PCR Fragment	Primer Sequence	Product Size (bp)	Mg ²⁺ (mM)	pH
exon 23	A 5'-CTGAAGTCAAATAATGAAGTCCCA (SEQ ID No. 132) B 5'-GTTTGCTTTCTGATATCTAAACACA (SEQ ID No. 133)	360	4.0	8.3
exon 24	A 5'-CTTGTGAGAGGCTATAAACTGG (SEQ ID No. 134) B 5'-GGTAAACAGTGTAGGAGTCTGC (SEQ ID No. 135)	267	1.5	8.3
exon 25	C 5'-GCTTGAAGGATGAGGCTCTGAG (SEQ ID No. 136) D 5'-TGTTTCAAGATGAGCAGCATGGG (SEQ ID No. 137)	461	1.5	8.3
exon 26	A 5'-CTTGTGAGAGGCTATAAACTGG (SEQ ID No. 138) B 5'-GGTAAACAGTGTAGGAGTCTGC (SEQ ID No. 139)	267	1.5	8.3
exon 27	A 5'-GCCATTTTCTCTTTAATTGGAAAGG (SEQ ID No. 140) B 5'-ATCTTATTCATCTTTCTGAGAATGG (SEQ ID No. 141)	274	1.5	8.3
exon 28	A 5'-TGAAATAGCCCAACATCTGACAG (SEQ ID No. 142) B 5'-GATTAATTTGACAGCTTGATTAGGC (SEQ ID No. 143)	291	1.5	8.3
exon 29	A 5'-TGAAATATAAACTCAGACTCTTAGC (SEQ ID No. 144) B 5'-GTACTGATTTGGAAAGACATCTC (SEQ ID No. 145)	303	1.5	8.3
exon 30	A 5'-GATGTGACAGTGGAGCTATGG (SEQ ID No. 146) B 5'-GGAAAAATGTGGTATCTGAAGCTC (SEQ ID No. 147)	307	1.5	8.3
exon 31	A 5'-AAGTGAGCAAATGTTGCTTCTGG (SEQ ID No. 148) B 5'-TCATTAGGAACTGAACATCAGC (SEQ ID No. 149)	304	1.5	8.3

PCR Fragment	Primer Sequence	Product Size (bp)	Mg ²⁺ (mM)	pH
exon 33	A 5'-TAAAGGATTAATGCTGTTAACAGTG (SEQ ID No. 152) B 5'-TCACACTGAGCATTTACTACCTG (SEQ ID No. 153)	360	1.5	8.3
exon 34	C 5'-GCAAAGGAAATGTAGCACATAGAG (SEQ ID No. 154) D 5'-AGGCTATAGGCATTTGAAAGAGG (SEQ ID No. 155)	491	1.5	8.3
exon 35	A 5'-GTAGGCTCCAGAAAGACCCAG (SEQ ID No. 156) B 5'-GAAAGGATGGGTGTGTATTCAGG (SEQ ID No. 157)	406	1.5	8.3
mutation 7	GD A 5'-ACAGGCCATAGTTTGCCAACCC (SEQ ID No. 158) GD D 5'-IGGTATTAGAATTTCCCTTTCTTCC (SEQ ID No. 159)	426	1.5	9.0
DJG RT-PCR	SEE 5'-TGAAAGAGAATATGGAAAGAGGCTTG (SEQ ID No. 160) E 5'-CTTTATGAAGCCAAATTTCTACCC (SEQ ID No. 161)	2002	1.5	8.3
P2934AT1	A 5'-TCAAAATCAGTCGCCTCATCCC (SEQ ID No. 162) B 5'-CAATGTATCAGTCAGGGTTCACC (SEQ ID No. 163)	168	2.0	8.3

The annealing temperature was 60° C for all primer sets.

Mutations are detected by amplifying *WRN* exons from genomic DNA and directly cycle-sequencing the PCR products by dye-terminator cycle sequencing (Perkin Elmer) and an ABI373 automated DNA sequencer. Prior to sequencing, the PCR-amplified exon fragments were purified using a QIAquick 8 PCR purification kit (Qiagen). The resulting sequences are aligned by FASTA analysis (GCG). Nucleotide differences between WS and controls are subsequently confirmed by sequencing the reverse strand

consequences of splice-junction mutations. RT-PCR products were synthesized from

mRNA isolated from lymphoblastoid cell lines (Qiagen Oligotex, Qiagen). The large genomic deletion was detected in genomic DNA using long-range PCR (Expand Long Template PCR System, Boehringer Mannheim).

Diagnostic Criteria. WS patients were from an International Registry of Werner's Syndrome subjects. Diagnostic criteria are based on the following signs and symptoms (Nakura et al. 1994). Cardinal signs are: 1) bilateral cataracts; 2) characteristic dermatological pathology (tight skin, atrophic skin, pigmentary alterations, ulceration, hyperkeratosis, regional subcutaneous atrophy) and characteristic facies ("bird" facies); 3) short stature; 4) paternal consanguinity (3rd cousin or greater) or affected sibling; 5) premature greying and/or thinning of scalp hair; 6) positive 24-hour urinary hyaluronic acid test, when available). Further criteria are: 1) diabetes mellitus; 2) hypogonadism (secondary sexual underdevelopment, diminished fertility, testicular or ovarian atrophy); 3) osteoporosis; 4) osteosclerosis of distal phalanges of fingers and/or toes (X-ray diagnosis); 5) soft tissue calcification; 6) evidence of premature atherosclerosis (e.g. history of myocardial infarction); 7) mesenchymal neoplasms, rare neoplasms or multiple neoplasms; 8) voice changes (high pitched, squeaky or hoarse voice); 9) flat feet. Diagnostic classifications are as follows: "Definite", all cardinal signs (#6 when available) and any 2 others; "Probable", the first 3 cardinal signs and any 2 others; "Possible", either cataracts or dermatological alterations and any 4 others; "Excluded", onset of signs and symptoms before adolescence (except short stature since current data on pre-adolescent growth patterns is inadequate) or a negative hyaluronic acid test. Family designations are as previously used (Nakura et al. 1994; Goddard et al. 1996; Yu et al. 1996).

Mutations in WS Subjects. Initial screening of the WRN gene was based on sequence from only the 3' end of the gene (exons 23-35). Thus the first 4 mutations (designated 1-4, Table 3) were in the region 3' to the helicase domains. In this mutation screening, primers amplify exons 2-35 along with approximately 80 bp of

screened for mutations. These subjects were selected based on haplotype analysis that

suggested that each subject might have a different mutation (Yu et al. 1994; Goddard et al. 1996). A total of 30 Japanese and 36 Caucasian subjects were ultimately screened for each mutation by DNA sequence analysis of the appropriate exon.

5

Table 6. Summary of WRN Mutations

Mutation	Codon	Exon	Type of Mutation	Nucleotide Sequence	Comment	Predicted Protein Length
none						1432
1	1165	30	substitution	CAG (Gln) to <u>T</u> AG (terminator)	nonsense	1164
2	1305	33	substitution	CGA (Arg) to <u>T</u> GA (terminator)	nonsense	1034
3	1230	32	4 bp deletion	gtag- <u>AC</u> AG to gt-AG	4 bp deletion at splice-donor site	1247
4	1047-1078	24	substitution	tag-GGT to tag-GGT	substitution at splice-donor site	1060
5	369	9	substitution	CGA (Arg) to <u>T</u> GA (terminator)	nonsense	368
6	889	22	substitution	CGA (Arg) to <u>T</u> GA (terminator)	nonsense	888
7	759-816	20	substitution	CAG-gta to CAG-tta	substitution at splice-receptor site	760
8	389	9	1 bp deletion	<u>A</u> GAG (Arg) to GAG (Glu)	frame-shift	391
9	697-942	19-23	deletion (> 15 kb)	-	genomic deletion	1186

Table 7.
Mutation Status of WS Subjects¹

Mutation	Japanese WS Subjects		Non-Japanese WS Subjects	
	Homozygous	Heterozygous	Homozygous	Heterozygous
1	SY ^D			
2	HH ^D , HM ^D , MH ^M , NN ^D		GAR ^D	
3			SYR ^I	
4	FJ ^D , FUW ^D , HA ^I , HW ^D , IU ^D , JO1 ^D , JO2 ^D , KAKU ^P , KY ^D , MCI ^D , MIE2 ^I , SK ^D , ST ^D , TH ^I , TK ^M , TO ^D , ZM ^D , 78-85 ^I			
5	KO ^D , OW ^P	KUN ^I	EKL ^D , AG0780 ^I , AG4103 ^M	DJG ^P , CP3 ^I , NF ^M
6			CTA ^D	SUG1 ^P
7	WKH ^D			
8				FES ^I
9				DJG ^P , SUG1 ^P

- 5 ¹The diagnostic classification is as previously described (Nakura et al 1994).
 Diagnosis categories: ^D Definite; ^P Probable; ^M Possible; ^I Insufficient data. The country of origin (ethnic
 group) of non-Japanese subjects are: AG00780, USA (Caucasian); AG04103, USA (Caucasian); CTA,
 England (India, East African, Asian); CP3, France (Caucasian); DJG, Germany (German); EKL,
 Switzerland (German); FES, Germany (German); NF, France (Caucasian); SUG, USA (Caucasian); SYR,
 10 Syria (Syrian). AG04103 and AG00780 were obtained as cell lines from the Aging Cell Repository
 (Camden, New Jersey).

Five new WS mutations were detected in the WRN gene (designated 5-9,
 Table 6). Two of the mutations (5 and 6) were single base substitutions creating
 15 nonsense codons. Mutation 5 results in a C→T transition changing an Arg to a

Caucasian subjects were homozygous, and 1 Japanese and 4 Caucasians were
 heterozygous for this mutation (Table 7). Mutation 6 is also a C→T transition changing

an Arg to a nonsense codon. One Caucasian WS subject was homozygous for this mutation, and a second was a compound heterozygote. The predicted protein product is 888 amino acids. A third substitution mutation (mutation 7) was a G→T change at a splice-receptor site, generating a truncated mRNA devoid of exon 20 and a prematurely terminated WRN protein at amino acid 760. A single Japanese WS subject was homozygous for this mutation.

Two deletions were observed. One (mutation 8) is a 1 bp deletion at codon 389 resulting in a frame shift and a predicted truncated protein 391 amino acids long. This mutation is found in one Caucasian patient as a heterozygote. The second (mutation 9) is a much larger deletion. This deletion was first observed in RT-PCR experiments when 2 different RT-PCR products were obtained from RNA prepared from subject DJG. RT-PCR products produced by primers 5EE and B (Table 5) yielded 2 different products, one with the expected size of 2009 bp, and a second, shorter product approximately 700 bp smaller. The DNA sequence of the shorter product revealed that exons 19 through 23 were missing. To further establish the nature of this mutation, primers (exon 18A and exon 24A, Table 5) derived from the exons flanking this potential gross deletion (exons 18 and 24) were used to amplify genomic DNA from subject DJG using a long-range PCR protocol. A single 5 kb fragment was observed corresponding to the shorter RT-PCR product. (The normal fragment, which is estimated to be > 20 kb was not observed.) The complete DNA sequence of this 5 kb fragment was determined and contained the expected 3' and 5' ends of exons 18 and 24, respectively. The exonic sequences were separated by intronic sequences adjacent to the 3' and 5' end of exons 18 and 24, respectively. No sequences from exons 19-23 were found in the 5 kb fragment. In other subjects and controls, the intronic sequence in the intron 3' to exon 18 contained 531 bp of unique sequence followed by a 241 bp Alu repeat element. Likewise, for the region 5' to exon 24, there is an Alu repeat element separated from exon 24 by 3,460 bp of unique sequence. The 4938 bp

recombination error at 2 highly homologous Alu elements within the WRN gene. A

primer set, GD-A and GD-D (Table 5) was designed to specifically amplify a short fragment (426 bp) across this junction point. A single additional Caucasian WS patient, SUG, was shown to contain this genomic deletion. Further PCR amplification of the exons within this deleted region demonstrated that both DJG and SUG are heterozygous for this mutation.

Origins of WRN Mutations. Because multiple subjects have the same mutation and because the same mutation was observed in different ethnic groups, at least some of the mutations likely originated in common founders. Evidence for a common founder was examined using 2 short tandem repeat polymorphisms (STRPs) within the WRN gene. These STRPs, D8S2162 and p2934AT1, were isolated from the same P1 clone (p2934) and are within 17.5 kb of each other. While D8S2162 is not particularly polymorphic (heterozygosity = 54% in Japanese and 70% in Caucasians) and is primarily a 2 allele system (140 and 142 bp alleles), p2934AT1 is highly polymorphic (heterozygosity = 78% in both Japanese and Caucasian populations). For mutation 4, which has only been observed in Japanese subjects, all but 1 subject had the D8S2164/p2934AT1 haplotype of 140-148 (Table 8). The single exception, JO2, has the haplotype 140-150, with the p2934AT1 allele being 2 bp different from the 148 bp allele observed in other subjects with mutation 4. This 2 bp difference may be the result of a 2bp mutation, as is commonly observed in dinucleotide repeat STRP loci (Weber and Wong, 1993). The haplotype data is consistent with a common Japanese founder and is consistent with the linkage disequilibrium observed in the same Japanese subjects for other markers in the WRN region (Yu et al. 1994; Goddard et al., 1996). For mutations 2 and 5, in the Japanese, the 896R18-p2934AT1 haplotypes for the small number of available subjects, are consistent with common founders for each mutation. However, the non-Japanese subjects with mutations 2 and 5 have discordant p2934AT1 genotypes when compared to Japanese subjects with the same mutations. These results do not support a common founder for both Japanese and non-Japanese subjects with

discordant for p2934AT1 (e.g. compare AG00780 to FK1). It should be noted that

absence of evidence for a common founder does not necessarily exclude the possibility of a single originating mutational event. Intragenic recombination and/or mutations creating new alleles at the 2 STRP loci could, over time, obscure the origins of the different WRN mutations.

Table 8. STRP Genotypes at the WRN gene¹.

Subject	Ethnic Group	Mutation	y896r18	p2934at1
FJ, FUW, HA, HW, JO1, KAKU, KY, MIE2, TO	Japanese	4	140/140	148/148
JO2	Japanese	4	140/140	150/150
HM, MH, NN,	Japanese	2	140/140	144/144
GAR	Hispanic	2	140/140	156/156
OW, KO	Japanese	5	140/140	148/148
AG00780	Caucasian	5	142/142	136/136
EKL, AG04103	Caucasian	5	142/142	128/128
CP3	Caucasian	5/?	142/150	128/142
KUN	Japanese	5/?	140/142	128/148
DJG	Caucasian	5/9	140/142	128/del ²

¹Genotype data for HH, SK, ST, TH, TK, and ZM was not available. For y896R18, alleles in bp (frequency for Caucasians, frequency for Japanese) were as follows: 136 (0.030, 0.025); 138 (0.020, 0.010); 140 (0.460, 0.576); 142 (0.337, 0.359); 144 (0.084, 0.010); 146 (0, 0.010); 148 (0.009, 0.010); 150 (0.059, 0). For p2934AT1, alleles in bp (Caucasian frequency, Japanese frequency) were as follows: 114 (0.006, 0); 122 (0, 0.009); 124 (0.011, 0); 128 (0.253, 0.079); 130 (0, 0.018); 132 (0.006, 0.009); 134 (0.046, 0.096); 136 (0.086, 0.009); 138 (0.011, 0); 140 (0.034, 0); 142 (0.052, 0.035); 144 (0.023, 0.061); 146 (0.023, 0.053); 148 (0.034, 0.132); 150 (0.034, 0.105); 152 (0.057, 0.123); 154 (0.063, 0.088); 156 (0.086, 0.070); 158 (0.098, 0.070); 160 (0.046, 0.018); 162 (0.029, 0.009); 166 (0, 0.009); 168 (0, 0.009).

The 5 mutations identified here demonstrate that WS mutations are not restricted to the 3' end of the gene, but are also found in other regions of WRN. In addition, mutations 5 and 7-9 each disrupt either part or all of the helicase region. Thus the WS subjects homozygous for this mutation will completely lack the WRN helicase

properly results in complete loss of an activity of the WRN protein. However, the WS phenotype in these subjects is not appreciably distinct from the WS phenotype

generated by the other mutations described here. Thus, all mutations in the WS gene may be complete loss of function mutations.

EXAMPLE 8

IDENTIFICATION OF MOUSE WRN GENE

The mouse WRN cDNA was isolated by screening a mouse splenocyte cDNA library at low stringency with human WRN cDNA as probe. The mouse cDNA sequence is presented in Figure 9. The homology between human and mouse WRN cDNA sequence is about 80%. On the amino acid level, the human and mouse WRN gene product show about 90% identity. Notably, the repeated exon in human WRN cDNA (exons 10 and 11) is only present once in mouse WRN cDNA.

Genomic mouse WRN clone was isolated by using mouse WRN specific primers to screen mouse genomic BAC library. The genomic DNA sequence is presented in Figure 6.

The genomic DNA sequence is presented in Figure 7 and SEQ ID NOS: 207-209. The DNA sequence is presented in Figure 6 and SEQ ID NOS: 205 and 206.

EXAMPLE 9

LOCALIZATION OF THE WRN GENE PRODUCT

A rabbit polyclonal antiserum raised to a peptide of WRN gene product is used in an indirect immunofluorescence assay to determine the intracellular localization of the WRN protein.

A rabbit polyclonal antiserum is raised to the peptide Phe-Pro-Gly-Ser-Glu-Glu-Ile-Cys-Ser-Ser-Ser-Lys-Arg (FPGSEEICSSSKR) (SEQ ID NO: 204) by standard methods (see Harlow and Lane, *Antibodies, A Laboratory Manual*, CSH Press, Cold Spring Harbor, 1989; *Current Protocols in Immunology*, Greene Publishing, 1995). The peptide corresponds to residues 1375 through 1387 of the WRN

fixed with 3% paraformaldehyde and permeabilized for 2 min with a buffer containing

0.5% Triton X-100, 10 mM PIPES, pH 6.8, 50 mM NaCl, 300 mM sucrose, and 3 mM $MgCl_2$ (see for example, Fey et al., *J. Biol. Chem.* 98: 1973, 1984). The cells are then stained for 20 min with a suitable dilution of the anti-peptide antibody (1:1500), washed, stained with a suitable second antibody (e.g., FITC-conjugated goat anti-rabbit antibody), washed, and mounted for visualization by fluorescence microscopy. Control stains include bis-benzimidine (Sigma, St. Louis, MO), which stains DNA, and phalloidin (Molecular Probes, OR, BODIPY 558/568 phalloidin), which stains filamentous actin.

As seen in Figure 9, the WRN gene product is almost entirely located in the nucleus. Nuclear staining is readily noted in the epithelial cells at the bottom left in panel A. These cells are close to the periphery of the expanding clone of human prostate epithelial cells. Cells that are not rapidly dividing (e.g., cells closer to the center of the clone), such as those seen in the upper right of panel A, are stained in both the cytoplasm and nucleus. The location and size of the nuclei in these cells is shown by staining DNA with the intercalating dye bis-benzimidine (Hoeschst 33258), panel B. The overall size of the cells and in some cases key cytoskeletal features are revealed by staining for F-actin as shown in panel C.

EXAMPLE 10

ISOLATION OF A PROTEIN THAT BINDS TO THE WRN GENE PRODUCT

A yeast 2-hybrid interaction screen (Hollenberg et al., *Mol. Cell Biol.* 13: 3813, 1995) is used to identify and isolate a cellular protein that binds to the carboxy-terminal 443 amino acids (residues 990 through 1432) of the WRN gene product.

A library of 1.1×10^6 independent cDNA clones generated from RNA isolated from stimulated human peripheral blood mononuclear cells is generated in pACT-2 (Clontech, Palo Alto, CA) that creates cDNA/GAL4 activation domain fusions is co-transfected into yeast containing pLEXA with the WRN gene fragment to generate

for histidine. 67 colonies grew on this medium. Of these, 60 were cured of the pLEXA

plasmid by growth on medium containing cycloheximide and mated with a yeast strain expressing a fusion of a "sticky" laminin and the GAL4 activation domain. 19 clones did not activate the sticky protein and underwent DNA sequence analysis. Of these, 6 contained sequences that did not match any sequence in GenBank by BLAST search.

- 5 Two other clones encoded carnitine palmitoyl transferase I and prolyl 4-hydroxylase B subunit. Six independent clones encoded a 70K component of the U1 snRNP complex (GenBank Accession No. M22636). Moreover, all six derived from the RNA recognition motif region of the 70K protein.

- 10 From the foregoing, it will be appreciated that, although specific embodiments of this invention have been described herein for the purposes of illustration, various modifications may be made without departing from the spirit and scope of the invention. Accordingly, the invention is not limited except by the appended claims.

Claims

We claim:

1. An isolated nucleic acid molecule encoding a WRN gene product.
2. An isolated nucleic acid molecule selected from the group consisting of:
 - (a) an isolated nucleic acid molecule as set forth in the Figures or complementary sequence thereof;
 - (b) an isolated nucleic acid molecule that specifically hybridizes to the nucleic acid molecule of (a) under conditions of high stringency; and
 - (c) an isolated nucleic acid that encodes a WRN gene product.
3. An expression vector, comprising a promoter operably linked to a nucleic acid molecule according to any one of claims 1 or 2.
4. The expression vector according to claim 3 wherein said promoter is selected from the group consisting of CMV I-E promoter, SV40 early promoter and MuLV LTR.
5. The expression vector according to claim 3 wherein said promoter is a tissue-specific promoter.
6. A viral vector capable of directing the expression of a nucleic acid molecule according to claims 1 or 2.
7. The viral vector according to claim 6 wherein said vector is selected

8. A host cell carrying a vector according to any one of claims 3 to 7.
9. The host cell according to claim 8 wherein said cell is selected from the group consisting of human cell, dog cell, monkey cell, rat cell and mouse cell.
10. An isolated protein comprising a WRN gene product.
11. An antibody which specifically binds to the protein according to claim 10.
12. The antibody according to claim 11 wherein said antibody is a monoclonal antibody.
13. The antibody according to claim 11 wherein said antibody is selected from the group consisting of an Fab fragment, an Fv fragment and a single chain antibody.
14. A hybridoma capable of producing an antibody according to claim 12.
15. A nucleic acid probe which is capable of specifically hybridizing to a WRN gene under conditions of high stringency.
16. A pair of primers capable of specifically amplifying all or a portion of a nucleic acid molecule according to any one claims 1 or 2.
17. A transgenic animal whose germ cells and somatic cells contain a WRN gene which is operably linked to a promoter effective for the expression of said gene, said gene being introduced into said mouse, or an ancestor of said mouse, at an embryonic

18. The transgenic animal according to claim 17 wherein the animal is selected from the group consisting of a mouse, a rat and a dog.

19. The transgenic animal according to claim 17 wherein WRN is expressed from a vector according to any one of claims 3 to 7.

20. An agonist of a WRN gene product.

21. An antagonist of a WRN gene product.

FIGURE 1

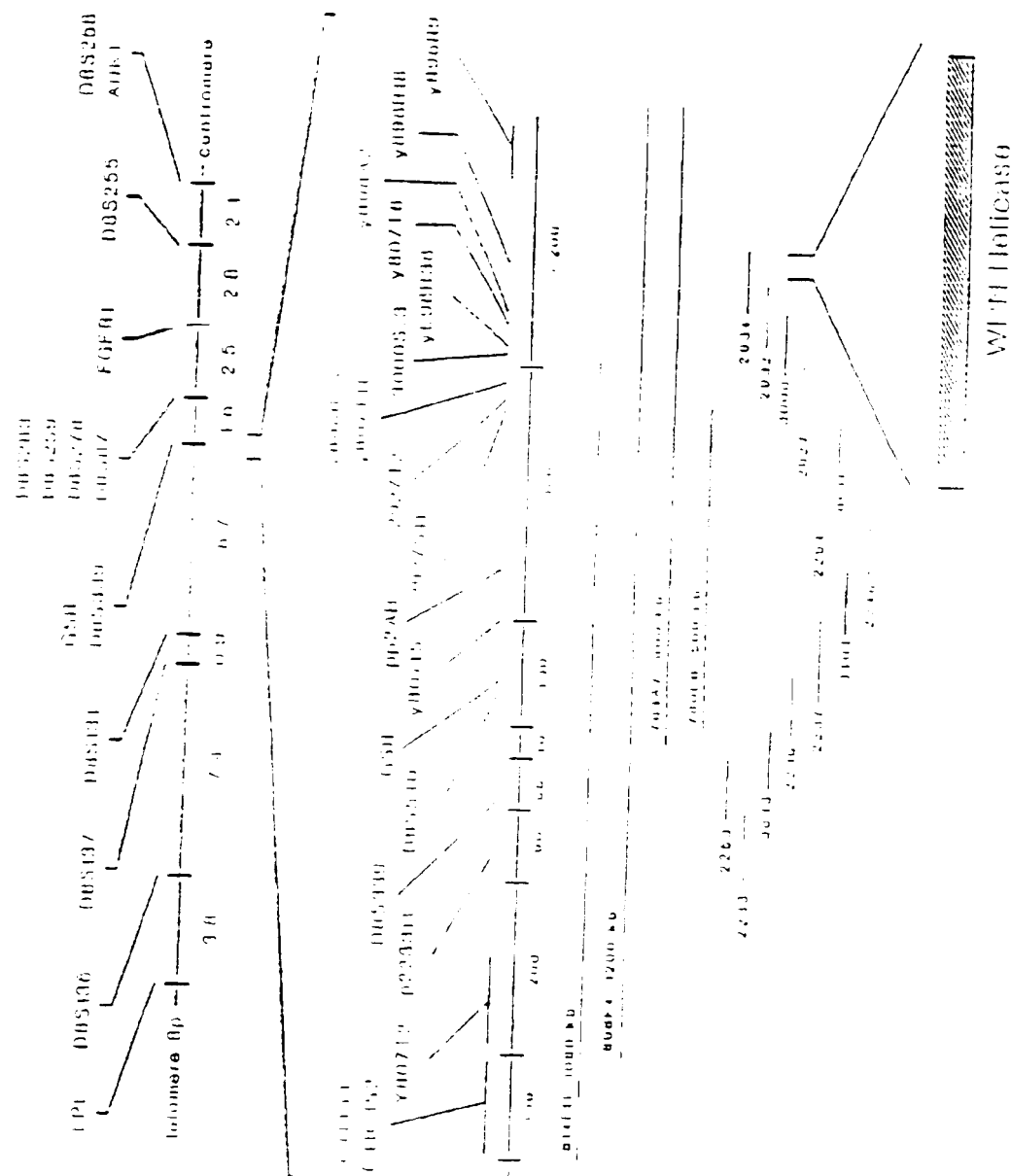


FIGURE 2A

[illegible]

1. The first step is to identify the problem. This involves understanding the current situation and what needs to be changed.

2. The second step is to set goals. These should be specific, measurable, achievable, relevant, and time-bound.

3. The third step is to develop a plan. This involves identifying the resources needed and the steps to be taken.

4. The fourth step is to implement the plan. This involves putting the plan into action and monitoring progress.

5. The fifth step is to evaluate the results. This involves comparing the actual results with the goals and making adjustments as needed.

6. The sixth step is to communicate the results. This involves sharing the results with the relevant stakeholders.

7. The seventh step is to review the process. This involves reflecting on the process and identifying areas for improvement.

8. The eighth step is to document the results. This involves recording the results and the process for future reference.

9. The ninth step is to celebrate success. This involves recognizing the achievements and the efforts of the team.

10. The tenth step is to learn from the experience. This involves reflecting on the experience and applying the lessons learned to future projects.

FIGURE 2A (CONT.)

TTACTTCTCTCCGAG JATCTAATTCTCTAGCCTTTTCTCCGATCAATATCCGAGGCGACAGTTTA	3240
TTTGGGACTGGGAGGRTCAACAGAGACTTGGTGGAGGCTTTTTCCTTTCTAGCTGATC	3300
ACTGAGGGATTTCTTGGTAGAAGTTTCTGGGTATACCAATTTATGAAGATTTTGGGCGCTT	3360
ACGAAAAGGGCTAGAAATTGGCTTCATAPAGCTATTACAGAACTTCAGAGGCTTCATGCTT	3420
CAAGCTAATGAAGATTTCTGTCCAAAGAACTTTCTTCTGCTAGTTTGGAAACTGTATCT	3480
TGGGGGACCCAAAGAGACTTTGTTATATCTAGTACCAGTTGAATTAAGTACAGAGAAAGAG	3520
TCTAACTTTGGAGAACTTATATTCTTTATTAACCATCTGATTAAGATTTCTTCTGGGAGTAAC	3600
ATTTCTAAAAAAGTATCATGGTAGAGTCAACCGAAGAAAGCTTACAGTTCTTGGAGGCTT	3660
GTTATTTTGGGACAGAGAGCAGGAGACTCAGATTCTGTTATATGGCAAACTGGTAGAAGCT	3720
AGGCGAAGACATGCCATATAAATGGATGTTCCCCCAGCTATTCTGGCAACAAACAGATA	3780
CTGGTGGATATGGGCAAAATGAGACCAACTACGGTTGAAAGCTAAAAAGGATTTGATGGT	3820
GTTTTCTGAAGGCAAGCTGCCATGTTGGGCGCTCTGTTGGGAGTCAACAAACATTTCTGC	3900
CAACAAATAGTGTTCAGACAGAGCTCTTTTCAATACAAACCTCAAGAAAGAAAGAAAG	3960
AGGAGTGTGTAGCAGAAAAATAAAAATGCTGCTTTCTGAGCTCTATGGGCACTGAGAAAC	4020
TCTTTATTCGAGAGAAAGAGATTCCTTTGAAGAGCTATAGCTGAGAGTACCAATCTCTGCT	4080
CTCATGCAATTTGGATGCACTTATCTGAGAGGCTGAAAGCTGCTGCTGCTGCTGCTGCTG	4140
GAGCTAGCAGGCTTCACTGAGAGCTTCAAGAGATTCATTTGCTGATGTTTACGCGAAGCTT	4200
CCCCCTCACTCAGCTATGATTAATAATTAGCTTAATCAGAACTGTTAGTTCTTGAAATATTT	4260
GAGAGCTAGCTTATGCAATGGCAATTCAGATCTTAAACATGCTCTGACAGGCGACTT	4320
CAAGCTTCATCTGATCTCAACAAAGAGAGATTTCTGCGGCTTCTGAAGAGATCTCTTCA	4380
AGTTCTAAG	4440
AG	4500
AG	4560
AG	4620
AG	4680
AG	4740
AG	4800
AG	4860
AG	4920
AG	4980
AG	5040
AG	5100
AG	5160
AG	5208

FIGURE 2B

MSEKMLSTTAQQRKQFENMIVQNKRCAYBERKQVRSKSTFEDDLPTLEPTGSIIVYSYDAS	60
DCSFLSEDIISMSLSDGGVVGTEMENPPLYNRGLGKVALDCLWSESNOVLPHYSSMSVF	120
PQGLHMLLENKQVKKAGVGLSDQDQWKLRLDFDTHLKNIVELTDVANKKIKOTETWSLNSL	180
VKHLGKQLLNDKSEIRCSNWSKEPLTEDQKLYRATDAYAGFTIYFNLEILDSTVQRFAIN	240
KEEEILLSDMKQLTGISSEVMDLAKHLPHAFSKLENFRVSIILKDISENLYSLRAMII	300
GSTNDETELRPNNLNLSSFEDSTTGGVQKQIPREHEVLEHVEDETWDPTLOHLAKHDGE	360
DVLGNHVERKEDGFEEDGVEDNKLKENMERACLMSSLDITTEHELQILEQQSQSEYLSDIAYK	420
STEHLSPNDNENDTSYVIESDDELEMEMLFWLSPNDNENDEYVIESDDELEMEMLKSLE	480
NLNSGTVEPTHSKOLKMERNLGLPTKEEEDDEENANESEEDDCKDFLNAPNNEEQVTCL	520
KQVFGHSESEKPVQWKVTHSVLEERRDNVAVMATGVGKSLQFVFPVTVGKIGLVISPLIS	580
LMEDQVLQKASNIPTACTLGSAGSENVLTQIKLGNVAVTVTPFYCSGIMGLLQQLERDI	600
GITLIAYDERHOCISENGHDFRDSERKLGLKLTALPMVPTVALTATASSSIRCDIVRCNL	660
RNPQTCTGFEFRANLYLEVRAKGNILQDLQPFVNTSSHWFEFTIIVOPSRMOTQQV	720
TGELRNTNLSDCTYKAGMSFTSTAKDTHHAFVDETCQVIAIAPENGINKVDIRQVINY	780
AFKHESEVYQETGRAGROGLSSSEHVLWAPADINLRAHLLTIPVINTPLNMLPDMARHE	840
KULHSEPOFFQILLRTEEDKQVTHASLSDMGTENCOONOFERLONOVEMDSEEDTSWDFG	900
PQAEVLLSANDILGENTFELFLFLPGENSLPLADQVSEHLEFOTONDQTESNWKATFR	960
QLITSEELIWEVRANKFHNICALTKHSPNNLHWNTESLILILQVIELDPKXFLPSSK	1020
TVSSGTNKHONVQVFNLSSTENHNLNVLNENHFCNLSSEENTSYVSDNTPGPEKAVSE	1080
SGPVDSAGQETQVLYGHLVFAEIKHAKHONVFPALLADYHILVDMARHAPFTTVENTH	1140
LDGVSEGNLAPLAPLEVTNHPDIONGVQDLFEETWSEEDQNTSLVANDHICTLSQEMV	1200
ITVSLFQEKHAPLKSIAESRILPLNTIGMHLSTAVYASQPLDPRAGLTPEVQKILRDVT	1260
RNEPWNEDMSHILIRMLVPEMIDTULIHMATETLWNGFDSGLQPSODVNRACFPSSSE	1320
LCGSHRSKEEYGINDETSSAEKARALPNTFANGETSEHLMONTKRGGLFS	1380
	1400

FIGURE 3A

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

FIGURE 3B

[illegible]

FIGURE 3C

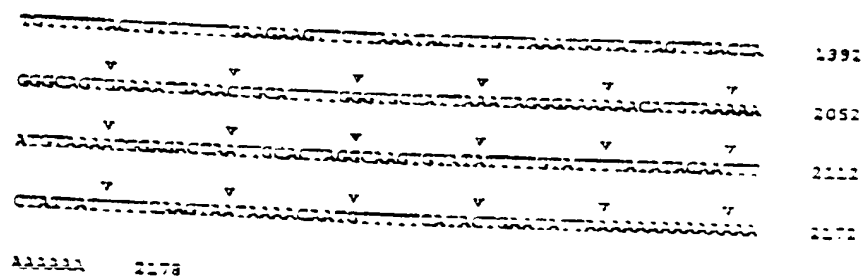


FIGURE 4B

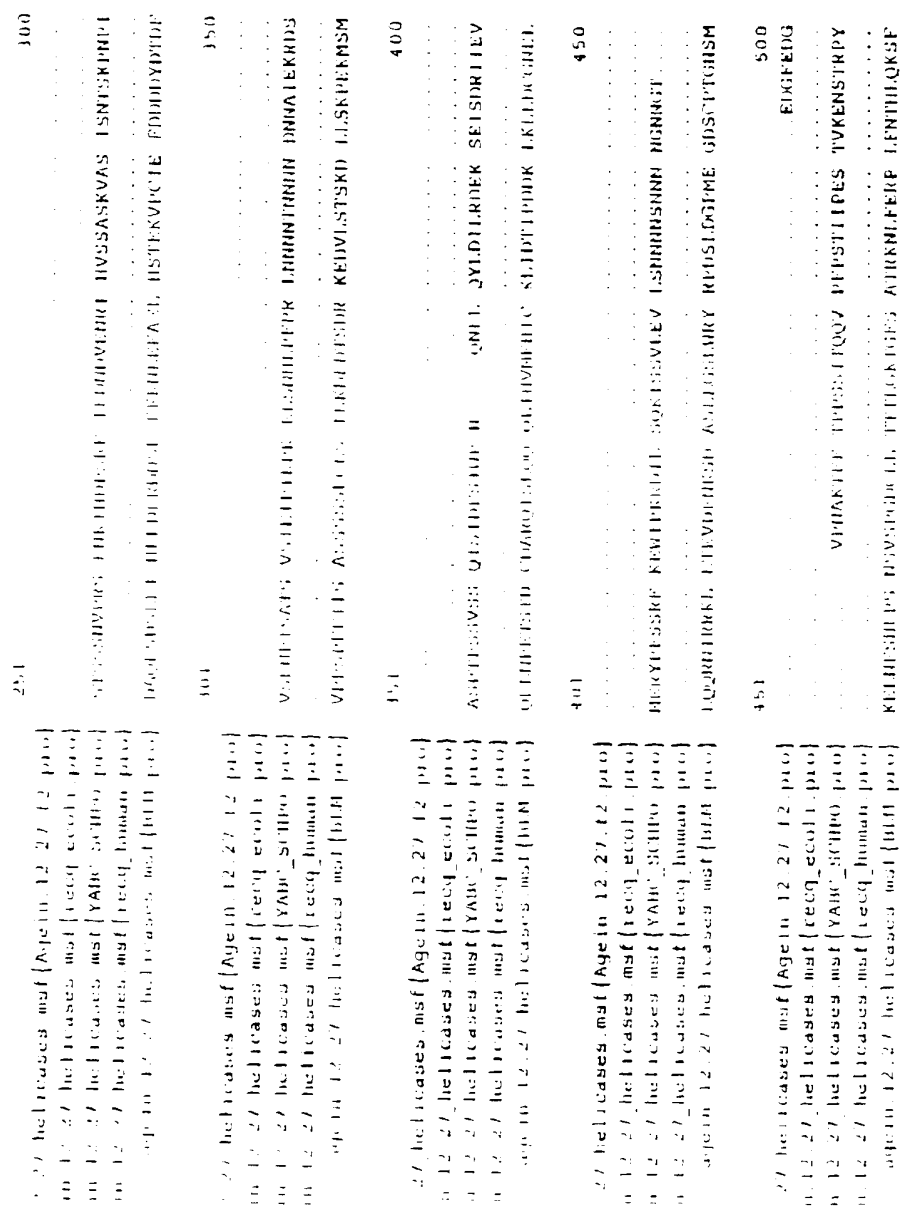


FIGURE 4C



FIGURE 4B

27 helices mat (Again 12 27 f2 pro)	1251	ESLHUKLTV YVSDDELSG VDEKSTFV QDPKAYSS QVLSAQQL	1400
12 27 helices mat (veg ecoli pro)	1251	ELHAKK	1400
12 27 helices mat (YAB SCIII pro)	1251	YVSDDELSG VDEKSTFV QDPKAYSS QVLSAQQL	1400
12 27 helices mat (veg human pro)	1251	ESLHUKLTV YVSDDELSG VDEKSTFV QDPKAYSS QVLSAQQL	1400
Again 12 27 helices mat (hlt pro)	1251	ESLHUKLTV YVSDDELSG VDEKSTFV QDPKAYSS QVLSAQQL	1400
27 helices mat (Again 12 27 f2 pro)	1401	EDVYVETV FAR QHQAQ DQVPAITAT HKILVIMAKM RPTIVEMK	1450
12 27 helices mat (veg ecoli pro)	1401	EDVYVETV FAR QHQAQ DQVPAITAT HKILVIMAKM RPTIVEMK	1450
12 27 helices mat (YAB SCIII pro)	1401	EDVYVETV FAR QHQAQ DQVPAITAT HKILVIMAKM RPTIVEMK	1450
12 27 helices mat (veg human pro)	1401	EDVYVETV FAR QHQAQ DQVPAITAT HKILVIMAKM RPTIVEMK	1450
Again 12 27 helices mat (hlt pro)	1401	EDVYVETV FAR QHQAQ DQVPAITAT HKILVIMAKM RPTIVEMK	1450
27 helices mat (Again 12 27 f2 pro)	1451	HWVSRKGA MIA PLTVI EHCQTHSVQ TDLFSSTPQ EQKTSIAK	1500
12 27 helices mat (veg ecoli pro)	1451	HWVSRKGA MIA PLTVI EHCQTHSVQ TDLFSSTPQ EQKTSIAK	1500
12 27 helices mat (YAB SCIII pro)	1451	HWVSRKGA MIA PLTVI EHCQTHSVQ TDLFSSTPQ EQKTSIAK	1500
12 27 helices mat (veg human pro)	1451	HWVSRKGA MIA PLTVI EHCQTHSVQ TDLFSSTPQ EQKTSIAK	1500
Again 12 27 helices mat (hlt pro)	1451	HWVSRKGA MIA PLTVI EHCQTHSVQ TDLFSSTPQ EQKTSIAK	1500
27 helices mat (Again 12 27 f2 pro)	1501	DKICTEQSM ATTSSTLQK KHLSTAES RILPMTGM HLSQAVKAG	1550
12 27 helices mat (veg ecoli pro)	1501	DKICTEQSM ATTSSTLQK KHLSTAES RILPMTGM HLSQAVKAG	1550
12 27 helices mat (YAB SCIII pro)	1501	DKICTEQSM ATTSSTLQK KHLSTAES RILPMTGM HLSQAVKAG	1550
12 27 helices mat (veg human pro)	1501	DKICTEQSM ATTSSTLQK KHLSTAES RILPMTGM HLSQAVKAG	1550
Again 12 27 helices mat (hlt pro)	1501	DKICTEQSM ATTSSTLQK KHLSTAES RILPMTGM HLSQAVKAG	1550
27 helices mat (Again 12 27 f2 pro)	1551	PLDERALF DEVKLTADV DRDPVSDM SKSLIRMLV PERIDTYLII	1600
12 27 helices mat (veg ecoli pro)	1551	PLDERALF DEVKLTADV DRDPVSDM SKSLIRMLV PERIDTYLII	1600
12 27 helices mat (YAB SCIII pro)	1551	PLDERALF DEVKLTADV DRDPVSDM SKSLIRMLV PERIDTYLII	1600
12 27 helices mat (veg human pro)	1551	PLDERALF DEVKLTADV DRDPVSDM SKSLIRMLV PERIDTYLII	1600
Again 12 27 helices mat (hlt pro)	1551	PLDERALF DEVKLTADV DRDPVSDM SKSLIRMLV PERIDTYLII	1600

FIGURE 4C

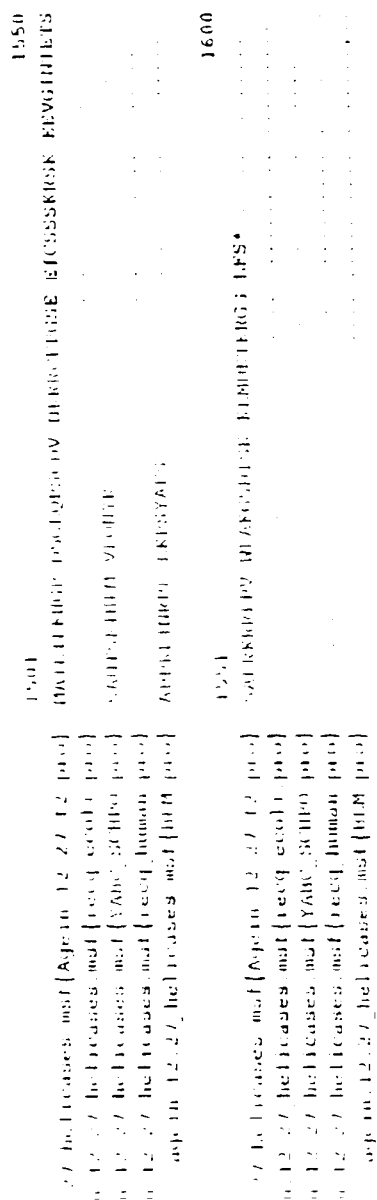


FIGURE 5A

TACATGAGGG	TTACTTTTCT	TTCTTATCTA	TTCTGATCTT	TACTGTTACT	AGCTTAAAAA	60
AAATTTTCTA	TTCTTTATCA	TTTATTTCTT	CTTAAAGCTT	CTTAAAGCTT	TTTAAAGCTT	120
TTCTGATCTA	TTCTGATCTA	TTGAGACATC	TATCTATTTT	AGCTTCTTAA	AGATGAGAGG	180
ATTAATGATC	ATGTTTCTCT	TTACCATCTT	TAATGAGAGG	CTTCTCTCTT	AGATTAATAT	240
TAGAGAACTG	TAATTAATCT	TTCAATCTGG	TAAGGACCTT	TTAAAGCTTT	TTCTAATGCA	300
AAAGTTTCTT	TTAAAGCATT	AAATATTTTG	AACTGCTCTT	CTTCTTTCTA	TTCTAAGGAG	360
TTCTGATCTA	CTACTTTTAT	TACTATTTCT	TATAGAGACT	CAAGCAAAAT	CTCTTTTCTA	420
CTCTGAGCTG	CTTAACTCTG	AAATCTGGCA	GATATCTCTT	AAATGATTTG	TTGAAATCTG	480
ATTGAACTCT	AGTTTCTCTG	AGGCACTTAA	CTTAAAGCAAT	AAATATTTCT	TACAAAAAGG	540
TATAAATCTA	AGCTTTCTAT	TTGCTTAAAT	ATGTTTCTCA	AACTGCTCTG	TTGATTTCTG	600
TTCTGATCTA	ATAGCTTACA	TAGCTTTTGG	AACTGGCATC	AAATAAGCTG	AGATTTCTCA	660
AUTGATCTCT	TACGCTCTCT	TTAAAGGACT	GGAACTCTAT	TATCTGGAGA	TTGCTAGAGC	720
TTCTGATCTA	GGACTCTTAA	CTTCTTCTCA	CTTCTCTCTG	CTTCTCTCTG	AGATTAATCT	780
AAATAGCTTA	AAAAATTTTA	TTTCTTTTAC	TTCTTCTCAT	TTCTTTCTTT	CTTCTCTCAT	840
AAAGCTTCTA	AGTTTCTCTG	CTTATTTCTA	CTTCTCTCTG	TTAACTTTCTA	CTTCTCTCTA	900
AGTAAAAAAA	AAAAAAAGAG	AAAAATTTAA	CTTCTCTCTA	TTCTAGGAGC	ATCTTAAAGG	960
AAAGCTTCTG	TTAAAGGCTG	CTACTTTGAG	ATGAGCTTAT	TTCTGCTCTG	CTTCTCTCTG	1020
CTTAAAGGCT	AACTTAACTA	GGACCTTCTT	TTGCTCTCTG	AGAGAGATCT	TATCTGATCT	1080
TTCTGATCTA	TTCTGATCTA	ATGAAAGCTT	TTCTTCTCTG	TTCTTCTCTG	CTTCTCTCTT	1140
GGCAATTTCT	CTAGCTTCTT	TACTTCTCTT	AGCTTCTCTT	TTCTTCTCTA	TTCTTCTCTT	1200
CTCTGATCTA	CTCTGATCTA	TTCTGATCTA	TTCTTCTCTG	TTCTTCTCTG	AGCTTCTCTA	1260
CTATTTCTCT	TTCTGATCTA	TTCTGATCTA	TTCTTCTCTG	TTCTTCTCTG	AGCTTCTCTA	1320
CTATTTCTCT	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1380
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1440
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1500
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1560
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1620
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1680
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1740
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1800
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1860
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1920
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	1980
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2040
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2100
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2160
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2220
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2280
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2340
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2400
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2460
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2520
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2580
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2640
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2700
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2760
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2820
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2880
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	2940
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3000
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3060
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3120
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3180
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3240
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3300
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3360
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3420
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3480
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3540
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3600
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3660
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3720
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3780
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3840
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3900
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	3960
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4020
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4080
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4140
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4200
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4260
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4320
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4380
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4440
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4500
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4560
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4620
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4680
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4740
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4800
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4860
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4920
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	4980
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5040
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5100
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5160
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5220
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5280
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5340
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5400
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5460
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5520
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5580
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5640
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5700
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5760
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5820
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5880
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	5940
CTCTGATCTA	CTCTGATCTA	CTCTGATCTA	CTCTTCTCTG	CTCTTCTCTG	CTCTTCTCTG	6000

FIGURE 5B

ATTACAGAG	TGTTGATGAA	TATGCTGCG	ATGGAGTTTA	AAAAAGAAC	AGTCTCTTAA	4250
TGTTGATGAA	TAATTTGAA	TAAATTTGAT	ATGGGAAAGA	AATGCTTAT	NATGATTTGA	4300
GAGAGGAACT	TAATTTGAA	ACTTGACTGG	TGTGCGATAA	TAACGCTTAA	TACTAAGAGA	4350
CACTTTTGGG	GGCGGAGAG	TCTTGAGAG	CTGATAGAGA	TAAGGCTTCA	NATTTGAGCT	4400
TGTTGATGAG	TGTTGATGCT	TAATGCTTAA	AGTTTCTCAT	ACAAATATAA	ATTAAGAGATA	4500
AGTCTCTTAA	TGGGAAAGG	TAAATTTTAA	TAAATTTGAG	AAGCATTTCT	TGAAGGCGAA	4550
AACTTCTCTA	CTGTAAATTT	GAATGAAAA	AGCTTAGAGT	TTTTCATAGT	TTTTTAAAGC	4600
AGCTTAAGAA	TTGATAGCTG	GACACAGGAA	GAATTTTCTT	TGAAAGGCAA	TTTTTATATA	4650
ATGCAAAAA	TGTTGATGCT	TTGTTGGGGG	CATTTGACTG	AAAGGGAATC	AACAGAACTT	4700
TGTTGATGCT	TAGAAATGCT	TTTATAGCTT	GATGGGGTGT	GGGTTTCTTA	GATAATGAAA	4800
AATAAACAGT	AAAAAATAAG	TAAAAAATAA	AGTAAGAAAG	TTGCGAATAC	AGTTTTCAT	4850
ATGCTCTCTA	TGTTGATGCT	GGACAGGGAG	CTTCTTACTG	AGATACCTTA	TGAGAACTTT	4900
CGATTATACA	AATTAAGAT	GATGGGAAAG	ATGGAAAAAT	ATCTTCTTCT	TAGCAGATTT	4950
AGGAGAGAGT	ATTTATATTT	TATTTTATGC	CAATAGTTATG	GAATTTATGA	TGATGCTTTT	5000
TTAAGAGAAC	AACTTGGCTA	AATAATATAT	AGTAATTTTA	AAAAATATTT	TGTTGCTGCT	5050
ACATGCTCTG	TGAATTTTAA	AGGCTAACTG	CTTCTCTCTT	GAATTAATCT	AACTCAATA	5100
ATGAAAGTGG	AACTTAATTA	ATTAAATGCT	ATCTTCTCTG	ATTAATTTTA	GGAACTGAAT	5150
TAACAGATAC	CTTCTCTCTG	TTTTATTTAA	TTTTAAAAAT	TGTTCTGGGT	AGAACTGCTT	5200
TGTTGATGCT	TGAGGAGCAA	CAAGTACAAA	AACTCTCTCT	GGGAATATAT	GGAACTGAAA	5250
AACTCTCTCT	TAATTTGAGG	TGAGGCTAAA	GAATTTCTTAC	TATAGATGGA	CAATTTAAAA	5300
GGGCTCTCT	CTTCTCTCTG	TGATCTCTTA	CTGAACTTTT	GTGAACTGAT	AACTATTTCA	5350
CTTCTCTCTA	CAAAACATAT	AACTCTCTTA	GATCTGAGAA	AGTAAATTAAT	ATTAACTGAT	5400
TTTCTCTCTA	AAATAGATTA	AGCAATATTA	TGTTAGCTCT	AAATTTAAAT	ATATCTATAC	5450
TTATCTCTTA	TGAAATGAT	CTAAATATAT	AGTTAAAGCT	GTAAATCTTA	AGAACTCTCT	5500
TAAAAAGT	AGCTCTCTCT	CTAGCTCTCT	CAAAAGAAAT	CACTCTCTAT	TATATAGATA	5550
CTTCTCTCTA	AGCACTTAAT	ATTAATTTCT	ATTAAGAAAT	TAATTAAGAG	GAATTTCTCT	5600
TTACTTTTAT	ATTAAGAAAT	CAAGCTTTCT	CACTAAAGAA	TATTTATCTG	ATTAAGAAAT	5650
ATTTCTCTCT	ATTAAGAAAT	ATTAAGAAAT	TAATTTATTA	CTGAGCTCTG	AGCAAGAAAT	5700
TAGCACTCTT	GAATTTCTCT	AGCAATATTA	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	5750
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	5800
TGATCTCTCT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	5850
TTCTCTCTCT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	5900
AGCACTCTCT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	5950
CTGCACTCTCT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6000
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6050
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6100
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6150
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6200
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6250
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6300
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6350
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6400
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6450
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6500
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6550
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6600
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6650
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6700
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6750
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6800
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6850
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6900
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	6950
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7000
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7050
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7100
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7150
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7200
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7250
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7300
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7350
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7400
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7450
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7500
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7550
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7600
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7650
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7700
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7750
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7800
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7850
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7900
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	7950
ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	ATTAAGAAAT	8000

FIGURE 5B shows a sequence alignment of the DNA sequence of the invention with the DNA sequence of the control. The alignment is shown in a table format with the sequence of the invention in the first column and the sequence of the control in the second column. The alignment is shown for a 1000 base pair region. The alignment is shown for a 1000 base pair region.

FIGURE 5D

TTTTACATAC	TTTGTATATC	TGAACTTTAT	TGATAGCTTC	TGAAAGCACT	TGCACATTTT	11660
AGATAATTTT	TATTTTCCCTA	AGTTTATATC	TAAATTTTTC	GAATTTTAAA	TATACATTTCC	11720
TTTGTATACG	TTTTTTTATC	TGTTTATATC	TTTATTTAGA	TTGATATTTT	TGTTATTTCTT	11780
GGATGACATC	GAGGATACAT	TTTGGGACTT	TTGTCGACAA	GCATTTTAAAG	TTTGTGTTTTC	11840
TTTGGACATC	TTAGGGCAAA	AAATTTGCAAT	TTGGGCTTCA	ATTTTATTTT	TTTGGAGGATC	11900
TTTGAATATA	TATTTTGTAA	TTTGGCTTAT	AGATTTTCTA	TTATCTTTTAT	TACGATTTCT	11960
TTTGAAGGTT	TTGAGTATTA	TGATTTTAAAC	TTTGTACTTA	GATGGGTGAA	TAGGAATCTCA	12020
TAGAGTCTTA	TTTAAATTTCA	TTTAAACACT	ACATCTCTTA	TTTGTGGATA	AACTGGTTTTC	12080
TATGGGCACT	TTTGTCTTTT	TGATTTTCTA	GCCTATTTTC	AGTTTCTGAG	CTCTAGTTTTC	12140
AAACATTTT	TTTAAATTTT	ACATTTTCTA	TTGGGTAAAT	TTTGAAGCTAG	TTTGTGCTTAG	12200
GAATATTTAG	AAAGATTTAC	TTTGAATTTA	TTTGAAGGTA	TTTAGAGCAA	AAATTTCTTT	12260
TTTGTGAGGA	AACTATTAAA	TTTTCGAGG	AAACATTTAG	TTTATAGGAT	TATTTTATTA	12320
ATTTTATTTT	TTTGTAAACA	TATTTTGGAC	ATAATTTTTC	TTTGTGCTGT	TTCTATGCTAG	12380
AAATATTAAA	AGTTATTAAG	AAATTTGTGT	TTTGTCTTTT	ACTAAATGAT	AAATTAATTA	12440
TTTAAATTTT	TTTCTTTTTC	TTTGGATGAT	TTTATTTTCT	GTATTTTCTT	AAACATCTAC	12500
GGTATACATC	TTTAAATTAAG	TAGGACCTTT	GCAGTTCTAG	AGCTTTTCTT	TTAGTCTTTT	12560
GGTTATATTT	ATTTTCTTTC	CTGATTTTGA	AGTTCTATTA	GGAAAGCTTA	TTATTTACCA	12620
CTGATTAATG	TTTAAATTAAT	CAAGAGCTTA	AGATAGCTTA	CTTTTCAAAA	AACTACATTT	12680
ATTTAAAAAT	TATTTTAAAA	TAGAGATTTA	AAATTTGCTA	TTTAAATTAG	ATTAAGGTAA	12740
CAATTTATTT	TTTAAATACG	GAATGCTTCA	TTTAAATTTT	TTTCTAGATT	TTTATTTTTC	12800
AAAGGAACAT	TTTGGAGCTT	GAGGCAAGGAG	AAATGGCTTA	ACCTTGGAGG	TTGGAGCTTTC	12860
AACTATTTTC	TTTGTCTTCA	TTTGTCTCTA	TTTGTCTCTA	GACAGCTTAA	TTTGTCTCTA	12920
ATTAATTTAA	AAATAGAAAG	ATATTTTCTA	AAATAGTTCT	GAATATTTCT	TATTAATATA	12980
TATTTGAAAT	TTTGTGCTTA	TTTGTGCTTA	TTTGTGCTTA	TACATTTTCA	ATGACATTTT	13040
AAATTTTCTT	TTTAAATTTT	TTTGTGCTTA	AGAAATTTTA	TATTAAGATCT	TTTGTGCTTA	13100
CAATTTTCTA	TTTGTGCTTA	TTTGTGCTTA	AGAAATTTTA	ATGAAATTTT	TTTGTGCTTA	13160
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13220
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13280
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13340
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13400
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13460
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13520
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13580
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13640
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13700
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13760
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13820
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13880
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	13940
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14000
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14060
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14120
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14180
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14240
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14300
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14360
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14420
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14480
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14540
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14600
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14660
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14720
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14780
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14840
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14900
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	14960
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15020
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15080
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15140
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15200
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15260
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15320
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15380
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15440
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15500
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15560
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15620
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15680
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15740
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15800
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15860
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15920
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	15980
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16040
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16100
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16160
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16220
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16280
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16340
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16400
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16460
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16520
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16580
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16640
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16700
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16760
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16820
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16880
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	16940
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17000
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17060
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17120
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17180
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17240
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17300
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17360
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17420
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17480
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17540
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17600
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17660
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17720
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17780
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17840
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17900
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	17960
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18020
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18080
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18140
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18200
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18260
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18320
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18380
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18440
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18500
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18560
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18620
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18680
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18740
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18800
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18860
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18920
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	18980
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19040
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19100
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19160
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19220
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19280
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19340
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19400
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19460
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19520
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19580
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19640
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19700
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19760
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19820
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19880
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	19940
TTTGTGCTTA	AAATAGAAAG	AGATTTTCTA	AAATTTTCTA	TTTGTGCTTA	AGATTTTCTA	20000

FIGURE 5F

AGTGAAGGAG	TTAAATTTGA	ATGATATGAG	TAAACTTTGG	AGTGTATGAT	GTGCAAAAGG	21120
ACAGTAAAGG	CAAAAATGTA	TTGGAAAGTG	TGAGTTGTTG	CACTGCACTG	ATGGATGAAT	21121
AGCTTTTGGT	TGCGATTTGG	TTTCTTTGTT	TTGTTTGTGT	TTGTTTTTTT	TAGATGGAGT	21122
CTTCTGATTT	TGCTGAGGCT	TGAGTGGAGT	GGGCTGATTT	GGGCTGATTT	TAGATGGAGT	21123
TTTCTGATTT	CACGCGCTGG	TTCTTGGCTG	GGCTTGGGAG	CAGCTTGGAG	CACAGTTTGG	21124
TTCTGAGGAC	TTTGGCTGAT	TTTCTGATTT	TTTCTGATTT	ACGGGCTTGG	ACGATTTTGG	21125
CGATGATGCT	TTTCTGATTT	TGAGTGGGTT	ATCTGCTTGG	TTTGGCTTGG	CAAGTGTGTA	21126
GGATTAGAGG	TATGCGGCTG	TTGCTTTTAT	ATATTTGCTG	TTGCTTGGAT	TTGAGCTGCT	21127
AGGATTTTGA	TGCTCTTTTG	TAAGCATGAT	TTGAGTTTGG	TAAGAGCTGT	AAATTAAGTA	21128
AGGTAAGTGT	CGAAACAGAT	ATGCAAGCTT	ATACACTTGG	CAATGCTTGT	CAATTTATAT	21129
TGAAATTTGG	TTGCTTTATG	TAAGACAGCA	AAGCAACAT	ATTTCTATTA	TGATTTATAT	21130
TGTGAGTGGG	GTGTTGAGAA	TGAGCAGGAT	GGGTATTAAG	TTTCTTAAAG	TTTCTTAAAG	21131
TGAAATTTGG	TTTAAATTTG	AAATATGAT	AGACAGTGT	GTGTTTAACT	TAAGTATATG	21132
GTTTAAATTA	CGATTTTATG	ATATCTGATG	TAATGCTTAT	TTGAGAAATG	TGAGCTTTGT	21133
CGGATTAAGT	ATGCGAGGCA	TAGTTGATTT	GGCTCTGGCA	AGGATCAAAAG	AGAGAGTTGG	21134
TGGAAGGCTG	TTTCTGCTTA	GGTGTATGCT	GAGGATTTGT	TTGTTAGAGT	TTCTGGGTAT	21135
TAAGAAATTT	TGAGAGTTTG	GGGCTTAAAG	AAAAAGGTAA	AGAGTGTAGG	AGTGTGGCTG	21136
TTTCACTTAA	TTTCTGATTT	CAGTGGAGAT	TAAAGATGCT	TTTCTGCTTA	TAATAGGCTA	21137
GAATTTGGCT	TTTAAAGGCT	TTTCTGCTTA	GGTCTTATTT	TGAGTGTGTA	GGTCTTATTT	21138
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	CAATTTATAT	CAATTTATAT	GGTCTTATAT	21139
AGTGAAGGAG	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21140
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21141
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21142
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21143
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21144
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21145
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21146
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21147
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21148
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21149
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21150
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21151
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21152
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21153
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21154
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21155
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21156
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21157
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21158
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21159
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21160
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21161
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21162
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21163
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21164
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21165
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21166
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21167
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21168
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21169
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21170
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21171
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21172
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21173
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21174
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21175
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21176
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21177
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21178
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21179
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21180
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21181
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21182
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21183
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21184
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21185
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21186
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21187
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21188
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21189
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21190
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21191
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21192
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21193
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21194
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21195
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21196
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21197
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21198
TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	TTTCTGCTTA	21199

FIGURE 5F is a sequence alignment of the DNA sequence of the invention with the DNA sequence of the reference sequence. The alignment shows that the DNA sequence of the invention is identical to the DNA sequence of the reference sequence. The alignment is shown in the following table:

FIGURE 5G

[illegible][illegible]

25268
25272
25280
25284
25300
25356
25440
25462
25668
25740
25800
25866
25920
25980
26040
26100
26156
26220
26280
26340
26400
26456
26520
26580
26640
26700
26756
26820
26880
26940
27000
27056
27120
27180
27240
27300
27356
27420
27480
27540
27600
27656
27720
27780
27840
27900
27956
28020
28080
28140
28200
28256
28320
28380
28440
28500
28556
28620
28680
28740
28800
28856
28920
28980
29040
29100
29156
29220
29280
29340
29400
29456
29520
29580
29640
29700
29756
29820
29880
29940
30000
30056
30120
30180
30240
30300
30356
30420
30480
30540
30600
30656
30720
30780
30840
30900
30956
31020
31080
31140
31200
31256
31320
31380
31440
31500
31556
31620
31680
31740
31800
31856
31920
31980
32040
32100
32156
32220
32280
32340
32400
32456
32520
32580
32640
32700
32756
32820
32880
32940
33000
33056
33120
33180
33240
33300
33356
33420
33480
33540
33600
33656
33720
33780
33840
33900
33956
34020
34080
34140
34200
34256
34320
34380
34440
34500
34556
34620
34680
34740
34800
34856
34920
34980
35040
35100
35156
35220
35280
35340
35400
35456
35520
35580
35640
35700
35756
35820
35880
35940
36000
36056
36120
36180
36240
36300
36356
36420
36480
36540
36600
36656
36720
36780
36840
36900
36956
37020
37080
37140
37200
37256
37320
37380
37440
37500
37556
37620
37680
37740
37800
37856
37920
37980
38040
38100
38156
38220
38280
38340
38400
38456
38520
38580
38640
38700
38756
38820
38880
38940
39000
39056
39120
39180
39240
39300
39356
39420
39480
39540
39600
39656
39720
39780
39840
39900
39956
40020
40080
40140
40200
40256
40320
40380
40440
40500
40556
40620
40680
40740
40800
40856
40920
40980
41040
41100
41156
41220
41280
41340
41400
41456
41520
41580
41640
41700
41756
41820
41880
41940
42000
42056
42120
42180
42240
42300
42356
42420
42480
42540
42600
42656
42720
42780
42840
42900
42956
43020
43080
43140
43200
43256
43320
43380
43440
43500
43556
43620
43680
43740
43800
43856
43920
43980
44040
44100
44156
44220
44280
44340
44400
44456
44520
44580
44640
44700
44756
44820
44880
44940
45000
45056
45120
45180
45240
45300
45356
45420
45480
45540
45600
45656
45720
45780
45840
45900
45956
46020
46080
46140
46200
46256
46320
46380
46440
46500
46556
46620
46680
46740
46800
46856
46920
46980
47040
47100
47156
47220
47280
47340
47400
47456
47520
47580
47640
47700
47756
47820
47880
47940
48000
48056
48120
48180
48240
48300
48356
48420
48480
48540
48600
48656
48720
48780
48840
48900
48956
49020
49080
49140
49200
49256
49320
49380
49440
49500
49556
49620
49680
49740
49800
49856
49920
49980
50040
50100
50156
50220
50280
50340
50400
50456
50520
50580
50640
50700
50756
50820
50880
50940
51000
51056
51120
51180
51240
51300
51356
51420
51480
51540
51600
51656
51720
51780
51840
51900
51956
52020
52080
52140
52200
52256
52320
52380
52440
52500
52556
52620
52680
52740
52800
52856
52920
52980
53040
53100
53156
53220
53280
53340
53400
53456
53520
53580
53640
53700
53756
53820
53880
53940
54000
54056
54120
54180
54240
54300
54356
54420
54480
54540
54600
54656
54720
54780
54840
54900
54956
55020
55080
55140
55200
55256
55320
55380
55440
55500
55556
55620
55680
55740
55800
55856
55920
55980
56040
56100
56156
56220
56280
56340
56400
56456
56520
56580
56640
56700
56756
56820
56880
56940
57000
57056
57120
57180
57240
57300
57356
57420
57480
57540
57600
57656
57720
57780
57840
57900
57956
58020
58080
58140
58200
58256
58320
58380
58440
58500
58556
58620
58680
58740
58800
58856
58920
58980
59040
59100
59156
59220
59280
59340
59400
59456
59520
59580
59640
59700
59756
59820
59880
59940
60000
60056
60120
60180
60240

	P_{in}	V_{in}	η
(a)		0	
(b)	100	100	1
(c)			0
(d)	8	100	
(e)		0	1

FIGURE 5H

[illegible][illegible]

FIGURE 51

[illegible][illegible]

FIGURE 5J

ATGTTGCTTC	TGATGAGCAT	TGGCATGCGC	TTATCGGAG	GGGTGAAAGC	TGGGTGCTTC	17360
CTGATTTTTC	AGGCGAGCGG	TCTGACTGCA	GAGGTTGAGA	AGATTATTTG	TGATTTTATC	17920
CGAAAGCTTC	CGGCGAGCTG	AGGTGAGAGG	GATGGGCTAG	CTCTGCGCTG	TTATGACTTT	17980
GATGAGCTTA	ACAGGCGATC	GACTATATTT	TTGACTGTTA	ACGCGATTTA	TGCTTTATGC	18040
TATTTGCTAA	AGCTTACCTT	TGTGATTTTC	TATGCTGCTT	TAGGAAAGCA	ATCTTTCTTC	18100
CGATTATGAC	TGCGAGGAAA	GTATGCTGAG	AAATTTTTTT	GTCTTTCTTC	GTCTTTCTGAG	18160
ACGAGCTTTT	TCTCTCTTTT	TTAGGCTGGA	GTGCGCTGGC	GTGATCTTGG	GTGGCTGCGA	18220
CTGCTATGTC	CGAGCTTGCA	GTGATTTCTT	TGGCTGAGCT	TGCTGAGTAG	GTGGGACTAG	18280
AGGCTTTCTC	GACGATGCTC	AGCTACTTTT	TGTATTTTTC	GATAGAGACA	GGGTTTCTCA	18340
TGTTGCTGTC	CGAGCTGCTG	AACTGCTGAC	TTCTGATGAT	CGGCGGCTGT	GAGCTTCTGA	18400
AACTGCTGCT	ATTAGAGGTC	TGAGCGATGG	GACCTGCGCA	ATACACTGAG	AAATTTTTAT	18460
TTTCTTTCTC	AGCTTAAGCT	TACGACTTTC	CGAGCATGCA	AAAGCTGCGC	CTGCTATTTT	18520
TTTCTTATTC	CTATCTGATC	ACTTGCAGAA	ACCTATTTTT	TCTGCGACTT	CACTTCTGCT	18580
AGCTTTCTGA	CTGCTAGTTC	TTCTCTAAAT	CTCTCTGAGT	CTCTCTGACT	GGTTTCTGCT	18640
GAGTAGCTTC	TCTAATGATC	ACAGTCTATC	GTATGCTTTA	CTTGTGACTG	GAAGGGACAG	18700
ACCAAGTTCT	TTGCGCGCTT	ACCTAGAGGG	ATTCTCTCTC	ACTTCTCTGC	AGAACTCTAG	18760
CTGCGCGAGA	CGAAGCGGCT	TGCTTGCGCT	TGTAGAAATA	TTTAAATTTA	TTATGCTTTT	18820
TTTTTTTAAAC	AGAAATTAAT	AGGAGATAGC	TTAGAGGATT	TTCTCTGCTA	GATTTCTTAA	18880
TACAAAGCTTC	GGGTCTTATA	ACTGATTAAT	TCTGATTAAT	TTCTTTTGAC	TCTTAGGATA	18940
GAGGAGTGGC	GATACGAATA	CGCTCATTTT	CAAGGCTGCA	GTGAAGATAG	TTTTTACTAC	19000
CTTAAAGTTC	TTGCGATTTG	TGAACTGCTT	TTGAACGATT	CTCTTAAAGA	ATTTGGAAGA	19060
TGCTTTCTTC	AAAGCGAGCT	CAATACTGTA	CTTAAATGCA	TTCTTCTGCA	AACTAGTCTC	19120
TTTAAAGATA	GGATTTATTT	TAGACTTTAG	TGCTGCTTTA	GTGAAGCTTC	AAAGAGTATA	19180
GAGTCTCTTC	AGCTTCTGCT	ATGAGGCTTC	GAAGCGAGCT	GAAGGAGCTA	GAGTCTGCTA	19240
AGTTGCGACT	TTGCTTATCT	AACTTCTGAG	GATGCGCGCT	AGTTGCGAGC	CGAGCTTCTC	19300
TAGACTTAAT	GAATTAATTA	TCTATTAATC	ATGCTGAGCT	GAATTTCTCA	ATACAAAGCT	19360
AAATTAATTA	TCTTCTGCTT	TAAGAGGCTC	GGCTTAAGCT	CGTCTCTATC	GATATTAAGA	19420
TGAGAGGCTC	TTGCTTCTTC	TGCTTCTGCT	CTGTAATCTC	AGCAATTTTC	GAGATTTAGC	19480
TGGAAGATTC	AGTTGCTGCT	AGCAATTTCA	GAGGAGGCTC	AGCAATTTTC	TCTCACTCTC	19540
ATTTTCTTAA	GAAGCTTAAT	AACTAGGCTC	AGGCTGCTTC	ATATGCTTTC	AGCTTCTGCT	19600
AGTTGCTTAA	TTGATTTCTT	AGGAGTCTTC	GAGCTCTGCA	TTTCTGCTTC	ATATTAAGCT	19660
GTGAGTCTTC	GAGTCTGCTT	GAGCTCTGCT	GAGAGCTTCA	GTGAGAGCTC	GTGCTGAGAG	19720
GAGAGTCTTC	AAAGATTTAA	ATTAAGCTTC	GGGAAGTCTT	GAGGAGGCTC	AAAGAGGCTT	19780
GGTTTAAAGC	GTGCGAGGCT	AGATGCTGCT	AAAAAGTCTC	TTTCTGCTTC	TATCTCTCTC	19840
ATTTGCTGTA	TGCTGCTGAT	TGCACTGCTA	AGGCTGAGCT	CTTCTGATAT	AGTAAGCTTA	19900
GAAAGTCTTC	GATCTGCTAG	CTGATTTTAA	AGTAAGAGCT	ATATTTTAAAT	GAATTTATTA	19960
TAATATGCTA	AAATTAAGCT	TTGAGATTTA	GAATTAAGAG	AGTAAGCTTA	TGCTATAGCT	20020
GAATTAAGCT	TTTCTGCTCA	AAATTTATTC	TTATTAAGAA	GAATTTCTTC	CGAGGAGCTT	20080
GTGCTGCTTC	TGATTTCTCT	ATAATCTTTC	GATTTTCTCA	CTTCTGAGCT	TGCTGAGCTC	20140
TTGATTTCTC	GTGCTGATTC	GAGAGCTTTC	TTGCTGAGCT	TGCTTAAAGC	GTGCTTCTAC	20200
TGAAATTAAT	GAATTTATTC	AGCACTGCTC	GTATGCTGCT	TTATTTCTAG	TTATCTGAGCA	20260
GAGATCTCTA	GTGATTTCTC	TTGAGCTGCT	GAGGCGGAGC	TTGCTGCTAG	CGGAGATGAT	20320
GTGATTTCTC	TTGAGCTGCT	GTGAGCTGCT	GAGACTTCTT	GTGAGAGGAT	GAAGAGGAGG	20380
GAAGCTTCTT	AGGCTTCTCA	GTCTTCTCTT	GTGAGGAGAT	GATAGGATTT	AGAACTGCTC	20440
AAATTTCTCA	GAATTTTATA	GAGATTAAGC	GTAAAGGAGG	GAAGAGGAGG	GAAGAGTATC	20500
TAATTTCTTC	TGAGGAGGAA	GAAGAGGAGG	TGAAAGGAGT	TAAGAGTTTC	GAAGAGTTTA	20560
TATTTCTAGC	TATTTCTAAG	AGGATTTCTT	GTGCTATATC	AGTTCTTCTG	TTTCTCTATTA	20620
TAACTTTCTC	TGAGATTTTC	TATTTCTAAG	ATGAGCTTTC	TATCTATGCT	AACTTTCTGTA	20680
TTTATTTCTC	TGATACTTTC	GAATTTCTCA	TGTAAGTTTA	TAACTTCTCA	TTTCTCTGAG	20740
GTGCTTCTTA	CGATTTCTCA	AGGTTTAAGT	AGGGAATTTT	TGAAGTAACT	TGTAAAGCTT	20800
CGAGGCTTTC	GTAGAGAGGC	ATTTTATTAAG	AAAGCTGCTC	AACTCTCTTC	GTCACTCTTC	20860
AACTTTCTTA	AGCTTCTGAA	TGAGGCTGCA	GAGAGGAGAA	TATTAATTTA	TTTCTTCTGAT	20920
TAGATTTCTC	GTCTTCTGAA	CTTCTCTCTC	TGCTCTGCA	ATTAAGATAG	CTTCTTCTAGC	20980
GAAGCTTAAAG	AGCTGCTGCT	CTTCTTCTAG	CTGATCTCTT	GGAACTTTTC	TTTCTTCTAGA	21040
AACTTATTTT	GTCTTCTGAG	TTAAAGGCTT	TTGCTCTCTT	GTATGAGGAA	TTGATTTCTTA	21100
GGGATTTCTC	TGCTTCTGAA	TTTCTCTGAG	TTGCTGAGC	TTGCTTCTGAG	CTTCTTCTTC	21160
AGGATTTCTC	GAGGCTGAGC	GGGCTGAGCT	ACTGAGGCTT	GTCTTCTGAG	AGGAGCTTTC	21220
CGAGATTTCT	GAAGCTGCTC	CTTCTTCTAA	ATATCTTAAAT	TTATCTGCTC	TTTCTTCTTC	21280
AGGCTTCTTA	TGCTTCTGAG	TTGCTGAGCT	GAGGCTGAGC	ATTTCTTCTA	AGGCTTCTTC	21340
AGGCTTCTAG	TTCTTCTGAG	CGAGATTTTC	GTCTTCTTAA	TATCTTCTTC	AGAACTGCTT	21400
GAATTTCTTC	TTTAAAGGAA	GAAGATTTTC	TTTCTTCTTC	TTTCTTCTTC	TTTCTTCTTC	21460

FIGURE 5K

CAAGGCTGTA	ATGCGAGCAG	TTTGGGAGGC	CGAGGCGAGC	AGATGACTTG	AGGTGAGGAG	42060
TTGAGAGACA	GGTGGGCGAA	CATGATGAAA	GTGGGTTTGT	ACTAAAACTA	CAAAAATTAG	42120
GTGGGCGTGG	TGTTGGGTGG	TTGTAAATGG	AGCTATTGAG	GAGGCTGAGG	CAGGAGAAAT	42180
GGTTGAAAGG	AGGAGGTTGA	GGTTGCACTG	AGTCAAGATT	GTGGCACTGG	ACTTTCAGCT	42240
GGGAGACAGA	GGGAGACTGT	GTGTGAAAAA	AAAAAAGAAA	AAAAAGAGCA	TAAAAAGGTA	42300
CACTGTGACA	GGGCACTTAC	CACGAATGGA	GGTGGCACTG	TGGGAGTTGG	TGTTGGGTTAG	42360
TGAGTTGAGG	AGGGGTGAGT	GAATGTGAAG	AGCTAGGACT	GTGGCACTGT	GTAGACTTTA	42420
TAAAGCGCTG	GCAGTTAGGC	CAGACTGAGC	GTGTGTATAC	GAGTGTAGCT	ACTGTATAGC	42480
GTAGCTGCAAT	ATGTACCTGT	CAAACTAAAA	CAAAAAGTTA	AAAAATTAGG	TGCTTTTGGG	42540
AATAATTAAT	TAACTTATGG	TTACTGTAAAT	GATTTTTCTT	TATGAATTAA	AATTTTTTTA	42600
GTGGGTGCGA	ATAAAGCTTG	GGTTAAAAAA	CAAACTATAT	GTACAGCTAT	ACAAATATAT	42660
TTGCTTTTAA	TGGGTTTCTT	TTAAGATTTT	TTGTGTTTTT	GATTTTGTAA	AATTTTGGTT	42720
TACTTTTCTG	ATTTTTTTTT	TTAAAAAGCA	AGACAAAAAC	CGACGACATA	GGTTAGGGCT	42780
ACATGGGCTG	AGGATGATGA	GTGTGACTAT	GTGGCACTGG	CACATCTTGT	GGACCGAGGT	42840
GTTCAGGGGG	AGTCATATGC	ATGGGGCTGT	CATCTCTTGT	GATAACAAAT	GTTTCTTTCT	42900
GACAGCTGCA	GAAGCGGCTG	GGTGTGTTAC	AGTGAATCTG	TAAAAAATTA	TAAAAATGAT	42960
AGTATAGCCA	AGACATAAAG	ATAATTAACAT	AGTCATTTAT	TATCATTTTT	AAGTATTAGA	43020
TACTGTACAT	GCTAGAGCTT	TACACAGCTG	GCAGCGAAGT	GAGTGTGTTT	GGTGTGTTT	43080
ACACGATTAC	CAAGCAAAAC	ACATGGGTGA	TGGTTTTGAT	TGTGATGTTA	CGATGGGCTG	43140
ATGTGACTAG	GTGGTAGGAA	CTTTTCACTG	GTATGATTAAT	GTAAATGGATA	TTTGTGCTGT	43200
TTGGGTGGGG	GTGGGTGAGT	GGACAGCTAT	TATGTGGTGG	ATGAGCTGTA	ATTAGATAGT	43260
GTTCAGAAAG	CTTTGGCACA	TTGGTAATAG	TAAATGGTGG	TGGCAATAT	GATGATGATG	43320
ATGATGATGA	TTGAGAGACT	AGATGGTTAA	ATTTTATGGT	GTGTGAAAGG	TACTGTGTAA	43380
AACTATGATA	TTGATAGAGT	TGTGCTTTTT	ATAGGGGCTG	TAAAGAACTA	TGAATTATAT	43440
TGAATAGGCT	TGTAAATGAG	TGTGTGTGCT	TGTGCACTGC	AGCAAGCAAG	GGTGTGCTAT	43500
CACATTAAGG	AACTTTTGGG	GATATGTTGG	TGATATGTTT	GAATGGCACT	TGTAAATGAT	43560
GTGATGATAT	TGTAGAGGGA	AAAGGTGAAG	AAAAATTTTA	TATGATGTTA	TGTAGGAGCA	43620
TATTTGAAGG	AGTGGTTTTA	TTTGGCTATG	GTAAAGAGTT	GAAGGCTGTT	TAAATTTGCT	43680
TTGAAATGAT	TTAGGGAGCA	AGTTATGTTA	TTTATAGATT	AGAAATGAAA	GTGGAAATAT	43740
TGACATTTTA	TTTGTGTTAT	TTGGAGCTTG	TTTTGCAAGG	AAAGACTGTT	AGGCTTAGAG	43800
AGTTATTTTA	TTTGTGTTAT	AGCATGTTTA	AGATTAATTA	GTGACAGACT	TGTGTGTTAT	43860
AGTGGCTGTA	AACTGGAGCT	GGCAATTTAG	TTAAAGAGAA	GAATATGCTA	ATTGCTTAAT	43920
GTGATATGAT	ATAGGGGTTT	ATAAACTTTA	AGTTGATCTG	TGAAAGAAAT	TAATGCTATT	43980
ATGATGATAT	TGATATGTTT	ATTATTTTAA	TGTGCTTAA	TAGGGGTTAG	TTTAAAGATA	44040
AGAAATATTA	AAATGAGGCA	GAATATTTTA	ATGAAAAAT	GTAAAGATAT	ACTTACTTAA	44100
GAATTTTGGG	TGCTTTTATA	CATGCTATAT	AGGTTGTACA	TGATATATTA	AAAGATATAT	44160
TTTGTGTTAT	TATTTGTTAA	TGTGAATAG	AGGATTTTAA	TAAGGTTGTT	ATTAAGCTAT	44220
TTTGTGTTAT	TTTGTGTTAT	GAGACAGGCT	GTATGTTGTT	TTTGTGAGCT	GGAGTTGGCT	44280
GGACAGTTGA	TAGCTTATCT	CAGGTTTCAA	GTGTTGGGCT	GAAGCAAGCT	TTGGCTTTTA	44340
GGTGTGTTAG	TAGGCTGGAG	CACAGTTTGG	TAAATTTTAA	GTATTTTCTT	AGAGATGAGT	44400
TGTGATGATA	TTGGCGAGCT	TAGTGTGTTA	GGGTTAGGCT	TAGGCGATGG	TGGGAGCTTA	44460
GGTGTGTTAA	GTGTTGGGAT	TACAGCTGTT	AGGCTGTTAA	TTGAGCGGCT	TTTGTATTTT	44520
TGTGATTTTA	CACTGCTAAT	CATTAAGCTA	GTATTTTGGG	ATGTCACACT	GTGTTGTTGG	44580
AGTGTGATAT	TTGCTTTATA	AAATAGCTTT	GGTGTGTTTA	TTGGAGAAAA	CATTTGTTGA	44640
TTTATTAATTA	TATTAAGAGG	AAATGAGGTA	TAGTATGTTT	AAATAGGATG	CAGATATGTT	44700
AGTGTGATAT	TTGATTTAAG	TGTAGGCAAG	CGCAATTAAT	TTAGAGGTTT	AAAGAGAGAG	44760
TTTGTATGAA	ATCTAAAGCT	TGTACAACTG	TGAATTTGAA	AGTGTGCTCT	ATTGCTATAT	44820
GATACAGATG	TGACCTAAAA	ATAGGTTAAT	TATGAAATTA	CAAAAGCTAG	GTATTAATTT	44880
CATAGCTGTA	TTGTTTAAAT	AGATTTTTGG	TGGAAAACTG	TTGATATTTA	AAAGAGCTTT	44940
GTAGCTGTTT	ATTGTTTTTT	TTTTTGGGTT	TTTATATTAAT	TTTAAAGCTG	GTATATGAGT	45000
TTTGGATGCT	AACTATGTTA	AAATTAATTA	TAAAGCAATG	NATGAACTAG	GTAGAGAGTA	45060
TTTGGATGCA	TTTATTTGGG	GGATTTTATA	TGTGCTGTTA	AGTGGAAAT	TAGAGGTGGG	45120
TTTGGATGTA	TTTATTTGAA	AACTAGGGTT	AAATGTTGTT	TTTTTAAAGG	AGTGAAGGCT	45180
GTAGAGGCTG	GTGTTGTTTA	ATTGTTGCTT	AGATATTTTG	AGGCTTTATG	TACATGCGAA	45240
AGGAGAGGAG	AAATAGATGG	AGGTTGTTAG	GTAAATTTGA	GTATTTGTTA	TATATATTAAT	45300
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45360
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45420
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45480
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45540
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45600
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45660
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45720
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45780
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45840
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45900
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	45960
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46020
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46080
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46140
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46200
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46260
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46320
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46380
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46440
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46500
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46560
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46620
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46680
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46740
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46800
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46860
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46920
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	46980
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47040
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47100
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47160
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47220
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47280
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47340
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47400
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47460
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47520
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47580
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47640
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47700
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47760
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47820
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47880
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	47940
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48000
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48060
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48120
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48180
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48240
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48300
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48360
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48420
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48480
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48540
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48600
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48660
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48720
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48780
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48840
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48900
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	48960
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49020
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49080
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49140
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49200
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49260
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49320
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49380
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49440
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49500
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49560
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49620
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49680
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49740
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49800
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49860
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49920
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	49980
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50040
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50100
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50160
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50220
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50280
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50340
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50400
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50460
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50520
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50580
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50640
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50700
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50760
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50820
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50880
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	50940
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	51000
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	51060
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	51120
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	51180
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	51240
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	51300
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	51360
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	51420
ATATGATGTT	TAATGCTGTA	TTAGAGGCTT	AAATTTGTTA	TTAGAAATTA	TTAGAAAAAT	514

FIGURE 5L

AAATTAATTC	GGATACATAC	ATACACACAC	ATACATACAC	ATTAAAAAT	GATGAATAG	46250
TAAATTTGG	GGATATTTCA	GGATATTTCA	ATGATATTAAT	GACCAAGAGA	CAATATACAT	46320
TTTATGTTGG	TTGAGAGATG	GGATCATACG	GTTCCTGATC	TTACTATGTA	TAGAGGTTAT	46380
CTTTGCTTTT	GTAGATTTT	GAATCTTTTA	TTAGTTTACG	CACTTACGTA	CTTTTATTTA	46440
GGCTGTATAC	TTTATTTTGG	TTTATTTTGT	TTTATTTTGT	TTTATTTTGT	GGTTTATTTT	46500
TGCAATATTT	TGAGGATGCT	TGCAATAGCA	AGCTAATTTT	TGAGGAGAGG	GAATATATTA	46560
AGCTGAAAAA	TATATATTTA	AAGAACTGCT	TTTATGTTTG	TGATTTGAAA	AATATCTTGA	46620
TTTCTGATAT	AGTTGATTTG	AAATGATGCT	CAATATTTTG	ATATGTTTTG	TAAATTAAGT	46680
GAAGATGAGT	GAATGATTTG	TAAATGCTTT	GGGTTTACCG	CACTGTGGAG	CAATGCAAGA	46740
GAATGATGAT	TTGAGAGGAA	GTCTGCAAA	ATTAATTTTG	GGTTGCTTAT	AGGAGGTTTG	46800
TACATTAATTT	TAGAGGTTGT	GAAGAAATTC	ACACATGCTG	AAATTAATTT	GTTCAGTTTG	46860
AAAAATGCTA	AAATTTCTTG	ATCTGCAAAA	TAGAGAAATAT	GTTCGCAAGG	ATCTGATTTA	46920
GTAAATGTTG	TTTAAAAATG	AAATGATGCT	GTTCATTTAT	CAATATTAAGG	CAATTAAGGA	46980
ATCTGATTTA	TAGAAATGCT	AAATTAAGGG	AGCAATTAAGT	TCTATTTTAA	TAAATGATTT	47040
TTTATCTTGA	TTTAAAGGCA	AAGGCTACCA	AGTGATCAGA	ATTAATTTAT	TTAAGAGTTA	47100
AACTTAATTT	AAATGATATG	TGGAATTTGA	AGCAATTTTG	GGTTTGAAGC	AAATTCAGCA	47160
ATGCTTTGTT	TGATTTTCTA	GGCAGCTCAG	AAAAAGACAT	TTGAAACACA	AAATTAAGAG	47220
TGCTTAATTT	AGCTAATTTA	GTTCGATGCT	TGGAGGCAAG	GGCTTCAAGT	GATACCTTGA	47280
TGGAGTTTGT	AGCTGATTTG	CAATGCTTAA	GGCTATAGGG	AAATTAAGGA	GGAAATTTTG	47340
CTTTGCGAAC	CTTCGATGCT	GGTTTCTAGA	GAATATTTTA	GTATTTCTAG	GTTCATAGTT	47400
TTTAAAAATA	TTTCTGTTAA	TTTCTGTTAA	AAATGAGTTT	TTTATGCTTT	TACAGATATG	47460
AGTAAATTTA	GGTAAATGAG	AAATGAGTTT	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	47520
AAATGATTTG	AGATGCTTAA	AGATGCTTAA	GAAGGCTTAA	TTTCTGTTAA	AGATGCTTAA	47580
GAAGAAATGA	GATGTTTCTG	GGTCTGTTAA	GAAGGCTTAA	GAAGGCTTAA	GAAGGCTTAA	47640
GAAGAAATGA	GAAGGCTTAA	GAAGGCTTAA	GAAGGCTTAA	GAAGGCTTAA	GAAGGCTTAA	47700
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	47760
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	47820
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	47880
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	47940
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48000
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48060
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48120
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48180
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48240
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48300
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48360
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48420
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48480
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48540
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48600
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48660
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48720
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48780
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48840
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48900
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	48960
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49020
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49080
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49140
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49200
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49260
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49320
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49380
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49440
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49500
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49560
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49620
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49680
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49740
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49800
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49860
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49920
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	49980
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50040
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50100
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50160
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50220
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50280
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50340
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50400
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50460
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50520
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50580
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50640
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50700
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50760
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50820
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50880
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	50940
TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	TTTCTGTTAA	51000

FIGURE 5M

ATGCTATATG	CAAGTAAATG	TGTGATGAG	TTGCTTTTGG	AGTGTATGCG	TGTGGGCTAT	50460
CTTGGGGTTC	TCTGTTTATG	AGACACCTTC	CTGCTGGGCT	CTGCTGGGCT	CTGTACAGAT	50520
GTCACTACTG	TGCAGGTCCT	AACTTCTGCT	GGTGGGTTGG	CTGTACATTC	TTACAGTTTT	50580
AGTTTGGGTA	GGATACCTTC	AAACTCTGCT	TTATTTGTAA	TGTGGACATG	GGATTTTGGG	50640
TTTCTACTAT	TAGTTTCTTC	TTAATCTCTG	AAATTGAGAA	GATTAAGAGG	TCTGTTCTTC	50700
CACTGCTGCT	CACTGCTGCT	CACTGCTGCT	TCTGCTATG	TCTGCTATG	TTTCTTATTT	50760
TTTCTGACTG	TATAGAGAGG	AACTATGAC	ATTTCTGCT	TGACCTGCTG	CTTGTGATAT	50820
AAATTAAGAT	ACACATATTT	TTATACAAAT	TGTTTTGTAG	AAAGTTTTAT	TACAGATGCT	50880
GATTTACAGG	TAAAAATGAC	TTATGAAAAA	AGTTTTGATG	ACAAATGATG	CAGGCTGGCT	50940
AACTAAATAT	ATGGATGAT	CTTGTGATTA	ATGTAATTA	AACTGTAAAT	TAACTTACAT	51000
ATTTCTGATG	TTGCTTACAT	CGGTATGATG	ACATATATAT	AGCAATGATG	TGTGCTGCTG	51060
CTATTTGCTAA	CTTAATGCTA	ATAGCTTGGT	AAACAGATTC	GGAGTATTTA	AAAGATGCTA	51120
AGAGCGGCTT	AAATTTTCTT	TCTTGGGAGT	TTTATGTTTC	TACTGATGAA	GGAAATAGAC	51180
AGTGGAGGCT	GTGTTTTCTA	TTAGGTAACT	TAGATATGAT	ACTGAGAGCT	TGAAATATCT	51240
ATTTCTGACA	CTCAAAAGAC	ACACTTATG	TAACTAAGCA	TTTCTGGGCT	TTTCTGATTT	51300
AAATTAAGAT	ATTATTAATA	TTGCTATATA	AACTGCTATG	ATGATATTTA	TAGAAATATA	51360
GAAGATAGAG	CTTAAGATGA	TATTTGCTAC	CAATTAATGA	GTTTGAGGAA	GAATTAAGGA	51420
TGTGTTTCTG	TATTTTACAT	TTATTTCTAT	TTAACTGCTA	AGAAATGCTG	GATGTTTACT	51480
AGTATTAATTT	CAATTTCTCT	GTGAAGAGCT	TGAAGCTTAA	GTAAAGCTGA	TAAATAAGCT	51540
ATACATTTAG	CAAGTGGCTG	AAATTAAGTT	CAAGCTGGCT	TGTGCTGGCT	TTTAAAGGCT	51600
GTGCTATTTG	ATGCTATTTAG	GTCTAGAGAT	GAATTAAGGCT	TCTGCTGCTG	CTTCTATCTG	51660
TGTTTCAATG	TACTGCTGCT	TTGCTGAGCT	GGGAAGGAT	TTTCTGCTAT	TATCTGATCT	51720
AGAAAGGCTG	AAATATGCTA	ATTTACAGAG	AACTTAAGCA	AAATTAAGCA	AAATTAAGCA	51780
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	TACATAGATG	TGAAAGAGAG	AAATTAAGCT	51840
AGCTTAAGAA	GAATATGCTA	AACTTCTGAT	ATGCTGCTAT	ATGCTGCTAT	TGAAAGAGCT	51900
CTTTCTGCTA	TGCTGCTGCT	TATGCTGCTAT	ATGCTGCTAT	TGCTGCTGCT	ATGCTGCTAT	51960
AGCTTAAGAA	GAATATGCTA	AACTTCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52020
TGCTGCTGCT	TATGCTGCTAT	ATGCTGCTAT	ATGCTGCTAT	ATGCTGCTAT	ATGCTGCTAT	52080
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52140
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52200
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52260
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52320
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52380
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52440
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52500
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52560
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52620
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52680
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52740
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52800
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52860
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52920
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	52980
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53040
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53100
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53160
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53220
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53280
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53340
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53400
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53460
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53520
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53580
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53640
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53700
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53760
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53820
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53880
GAAGATTAAG	AACTTCTGTA	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	53940
ATGCTGCTAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	GAAGCTGAT	54000

FIGURE 50

TAGTTTCTGA	TTTGTATAGA	TTTGAAGTCT	TTTATGGAAG	AGAAATTCGA	AACGCTGGA	58860
ACAGTTAAGA	AGGTTAGTGA	TAGGCTTGAT	GGGCTAGCTG	AGTTGAGGAG	TGTGAGCTAT	58920
GTGAAAAAGT	AGTTAGCAGA	TTTGGTTGG	GAACATTAAT	GCTAAATGAT	TTGAGATTAAT	58980
GTACTCTCTT	CTTTACTGGG	TAAATCTCTT	TAAACCTTGA	GTAAATTTTA	GACATAAGTA	59040
GTAAATATNG	GATAAATTA	AGCTTTGCTA	GTATCTCTTT	AGGATTAAGG	ACTAATAAGT	59100
ACATATTTGA	TATTTAAGCA	TTTGTAACTG	TTGAGATTAAT	TTATCTTACT	CAAGTAACAG	59160
ATTAATCTTG	TGACTCTGAT	TTAAAAATTA	TGATTTAAGA	ATTAGTTAGT	GTCTCTGAT	59220
TTTAAATAGA	AAGCTTGGCA	TTTGAAGGCT	ATTCTGCTTT	ATTATTTGAG	ATATTTCTCA	59280
TGAATCTTGA	AGCTTTCTAT	TGCTAGAGAG	TATTTAAGCA	GATTTTACTT	GGAGAGGAAA	59340
GATATCTTAAT	AGGTTACTTT	TAACTCACTAC	TAAATTACTA	CAATTACTAT	ACTCTATTAA	59400
TATCTGATTA	TTAAAACTTA	GAGCTCTTAAT	TATCTCTGAG	AAAAATTAA	AAACTTTAGC	59460
CTCATTAATTA	CTTTTACTTT	CTAACTGAT	CTTTTAAAAA	TGACAAATAT	GTATTATACT	59520
TATTTATCTT	TTTCTTCTAGT	AAACATTTTC	ATTCTTAGCA	TGCAAGACCA	CACTCTAGAC	59580
ACAGCAAAAG	TGGAATAAAA	GGAAAGAAAT	GGACACACAG	TCTCATGCTT	AAGAGGGACA	59640
GATTTACTCT	GAAGATTTGA	TGAAAAAACA	TGCAACACCA	ACTTTTCTAC	AAGAAACAAA	59700
ACATTTTAAA	GAAGACATTT	ACTTCAAGCG	GGCGGGCTGG	CTTACGCTTG	TAACTCTAGC	59760
ACTTTTGGAG	GGCGAGCTGG	GTGCTACAGG	AGGTCAGGAG	TTGCAAAACA	GACTTGGGAG	59820
TATGCTGAAA	TTTCTCTCTT	ACTAAAAATA	CAAAAAATAG	CGTGGGCTGG	TGCTGCTGCG	59880
CTTCTGATCG	AGCTTACTAG	GAGGCTGAGG	CAGAAATATC	GCTTTGACCT	GGGAGGGAGA	59940
GTCTGCACTG	AGCTGAGATG	AGGCTTCTAG	GCTGAGGCTT	GGGCAACAGA	GGGAGACTCT	60000
GACTCTAAAA	AAAAAAAAGG	AAAAAAAACA	AGAAAACTTT	TACTTCTACT	AATAAGATAT	60060
GAGATTAATAT	TGACTCTCTG	ATGAAAAAAA	AGAGGACATG	ATGTTAAAAA	TTTCTCTCTT	60120
TTTCTCTCTT	AAATTTATAG	ACTGAAAGAG	TGCTTAATCA	TAACTTTCTT	TATTTTACGG	60180
GAATTTAAAT	TGAAAGCTGG	GAAGCTCTAT	TGAGAGCTTT	ATGATCTGCA	GGTCTTTTGA	60240
ATGATATATG	ATGCTGATTA	TTCTCTTACT	TAAAACTTGA	AGAACTTTAT	AAAAATAGCA	60300
GAATCTATAT	TTCTCTCTTA	TGCTGCTCTG	TCTTAACTCT	TTGCACTTAA	GTGATCTCTG	60360
GAGCTGATAT	AGAACTCTGG	GAJGCTCTAG	TGCTTAAGCT	TGTATTAATG	CTCTTAATAT	60420
GAATTTATAT	TGCTATTAAT	ATTTTAACTT	TGAAAGACTT	TTTAAAAAAA	AGGATCTGGA	60480
TCTCTCTATG	TTTCTCTAGG	TGCTCTCTAG	TTCTTAAGAT	GAAGTCTATG	TGCTCTCTCT	60540
AGCTTTTAAA	GAJGCTCTAT	TAGAGCTCTG	AGCTGATCTT	GGGAGCTAT	TTTCTCTCTT	60600
AGCTGATATG	AGAGCTCTGG	TGCTGCTCTG	AGTAACTCTA	AGTCTGCTCT	TGCTGATCTA	60660
GCTGATCTGA	TTTCTGAACT	CTGCTGCTCT	AGGAATCTAT	TAGGCTTTAT	CTTTGAGTAT	60720
TGAGATCTCA	GGGATCTCTG	TGCTGACCTA	GCTTAATATAT	TTGCAAGATAT	GGGCTTTCTG	60780
TCTCTCTCTG	GGAGATCTCT	GAATCTCTCT	TGCTGAGCTT	GGGCAAGTAT	TGGGATCTCA	60840
TGCTATCTCT	GGAGATCTCT	TGCTGAGCTT	TGAAAAATAT	ATGATCTCTG	AGGCTTTCTG	60900
GGATCTCTCT	GAAGATCTCT	ATTTCTCTCT	TGCTGATCTT	GAATCTCTCT	CTGCTATCTG	60960
CTCTCTCTCT	CTATCTCTCT	ATGCTCTCTA	TTGCTGCTCT	AGGAGGATCT	TAGGCTCTCT	61020
CTCTGATCTG	TGAAAACTCT	TAGGCTCTCT	ATATGAACTT	GAAGGATCTT	GAAGGATCTT	61080
TTGGAATCTA	AGGGAATCTT	TGCTCTCTCT	TTGCTGAGAT	ATGCTGAGAT	TGCTTCTCTG	61140
TGAAAGATAT	GAAGGAACTG	TGATATCTCT	GAAGGATCTT	TTTCTGAGAT	TGAGGCTCTG	61200
AGAGCTCTAT	GAGAGGATCT	GTGCTCTCTG	GTGCTCTCTG	GAGGAGCTCT	AGTACCTCTAT	61260
TGCTCTCTAG	TGCTGAGCTG	GAAGGAACTG	GTGCTCTCTG	GTGCTCTCTG	GTGCTCTCTA	61320
GATGAGATAT	GTGCTCTCTG	AGGGAATCTG	GTGCTCTCTG	GAGAGCTCTG	TGCTCTCTCT	61380
AGAGCTCTCT	TGAGATCTAG	AGGAGCTCTG	AAATCTCTCT	GGGAGCTCTG	GTGAGCTCTG	61440
AGGCTCTCTA	GTGCTCTCTG	GTGCTCTCTG	GTGCTCTCTG	GAGCTCTCTG	GTGCTCTCTA	61500
AGAGATCTCT	GAATCTCTAG	GAAGGCTCTAT	GAAGGCTCTAT	GAATCTCTAG	ATTAAGAACT	61560
AGAGGAGGAG	TAGCTCTCTA	TGAGGCTCTG	TGAGGCTCTA	AGAGGAGGCT	GTGCTCTCTG	61620
TAACTGAGGAG	AGGAGGCTAG	GTGAGCTCTG	TAGAGGCTCT	GTATCTCTAT	GAATAGAGAG	61680
AGGAGGAGG	AGGAGGCTAT	GATGCTCTAG	GTGCTCTCTG	GATGAGGCT	TGATATCTAT	61740
GCTAAAGGAG	TTGAGATCTG	GTGCTCTCTA	AGAAAGCTCT	GGGCTCTCTG	ATGAGCTCTG	61800
TGAGGATCTG	TTTCTGCTCT	GAGATCTCTG	TGCTGAGCTT	GTGAGAGCT	TGCTCTCTAT	61860
AGAGGATCTG	GCTGCTCTAT	TAGAGGCTCT	GTGAGGCTCT	ACTGAGCTCT	TGCTGAGCTT	61920
GAGCTCTCTG	GTGCTCTCTA	AAAGAGCTCT	AAAGAGCTCT	GTGCTCTCTA	GTGCTCTCTA	61980
AAAGCTCTCT	GTGAGGCTCT	AGTCTCTCTA	TGCTGAGGAG	GTGCTCTCTG	GTGAGGCTCT	62040
AAAGCTCTCT	AGGAGGCTCT	GTGCTCTCTG	ATGCTCTCTG	GTGCTCTCTA	TGCTGAGGAG	62100
AGAGCTCTCT	GTGAGGCTCT	GTGCTCTCTG	TGAAAGCTCT	GTGCTCTCTG	GTGCTCTCTG	62160
TTCTCTCTCT	AGGAGGCTCT	ATGCTCTCTA	AAATCTCTCT	GTGCTCTCTA	TTTATCTCTG	62220
TGAGGATCTG	GTGCTCTCTG	GTGCTCTCTG	AAAAATCTCT	TAGAGGCTCTG	GTGCTCTCTG	62280
GCTCTCTCTG	GTGCTCTCTA	TGAGGATCTT	TAGCTCTCTCT	TTAAATCTCT	GTGCTCTCTG	62340

58860
 58920
 58980
 59040
 59100
 59160
 59220
 59280
 59340
 59400
 59460
 59520
 59580
 59640
 59700
 59760
 59820
 59880
 59940
 60000
 60060
 60120
 60180
 60240
 60300
 60360
 60420
 60480
 60540
 60600
 60660
 60720
 60780
 60840
 60900
 60960
 61020
 61080
 61140
 61200
 61260
 61320
 61380
 61440
 61500
 61560
 61620
 61680
 61740
 61800
 61860
 61920
 61980
 62040
 62100
 62160
 62220
 62280
 62340

FIGURE 5P

[illegible]

FIGURE 5R

[illegible]

FIGURE 5S

GGGCTACGAA	GTGAGACTGT	GTTCGAAAA	ATGAAAAATA	TATATATAT	GATATGCGAA	75660
GTGCTATAGG	AGCTAGAGTG	TAAAAATGAG	ATGAGACAGG	TGGGATTGTC	TGTTTCTCTT	75720
GTTCAGCGGT	AAGCTGCGAG	ATTATATTTG	ATGATGAGGT	GTAGCGTGGG	AGTCCGTTAG	75780
GTTCAGAAATG	GTTCGTGTGA	AGGTTCTCTT	GTTCGTAAAA	TAGGAATGTC	AGGAAAAAGT	75840
TGTTGTCTTG	GTTCGTATTT	TGGTGTGGGG	GGATGTATAT	TAGGCTTTTA	GCTATTTGAG	75900
GTATTTGGGA	AAGTTGAAAG	GAGGAGCTGA	GTTCGAGCTG	GTAGAAATATA	GAATATGAAA	75960
GGATTTAGAT	GTTCATGTGG	TTCAGAGGCT	GGGCTTCTCT	GGGACTTTCT	TGTTGTGTGT	76020
GAGTTTACAA	GTTCCTCTTA	TGTTTAACTT	GGTTTATGTG	GTTCCTTACT	AGTTGTAGCT	76080
AAAACTTTTA	TTCGCGCATG	ATGTACTTTA	TGTTATCTCT	GTTCCTTCTT	TATTACAGCT	76140
ATATTTACAA	GTTCATTTCT	ATTGAGCTGA	TGATTTTAA	GAGGCGGAGC	AACAGGCTGA	76200
GTGCTTAACT	TGCTGTGAGT	TGTTTAACTG	CGAGGCGCTG	TGTTTGTCTT	TGAAATCTTA	76260
GTGCTTCTTG	TAGCTGTGAG	ATACTATTTT	TGTTAAAAAT	TATTTTCTTT	TGAGGCGAGG	76320
GATCTTCTCT	TGTTGCGCGG	GCTGCGATGG	AGTGGGATTA	TGATAGCTTA	GTTCAGGCTG	76380
AAAGCTCTTG	GTTCGAGTGA	TGTTTCTCTT	TGAGTTTCTG	TAAAGTCTGG	GATTACAGCT	76440
GTGAGTAACT	GGGCTGTCTG	TATTATTAT	GATTCGAGAT	TACAGATGAG	GAATTAAGG	76500
GTTTAGGAGG	GTTCATTAAT	TGTTAGATTT	GTTCATTTAGT	TAGCGGCGAG	GCGAGGATTT	76560
AAAGCTTACG	GTTCAGGACT	GTTCAGACTAG	TGCTTGGCAG	TGTGATATGG	GCTTTTCACT	76620
GTCTTCTCTT	ATGCTGTGATG	ATGATATCTT	TGTTTCTCTG	GTTCCTGGGA	AGTCTTTTAA	76680
GTTCGAGCTG	GTATTTCTTG	GAGGATTTCT	GTCTGATATA	TAACTCTCTT	ATTACTCTCT	76740
TGCTTTCTTT	TGTTTATGAG	GTACAACTTT	TTTATTTCTT	TAAAGCTATTT	GCTCTTATAT	76800
GTTAGATGCA	AGCTCTCTTT	GATTTATAGT	GTTCATGCGA	TAGGCTATTT	TGAGGCTTTG	76860
TGTATCTTTT	GTTCAGAAAA	GTTCATCTTA	AGCGGCGATAT	ATACTTCTCT	TGCTTCTGCT	76920
TGTACTCTTT	TTCAGCGATA	TATAGGCTTG	TGTTTCTCTT	TAGCTTATTT	TGAGGCTTAT	76980
ATATCTCTTT	TGTAGAGCTT	GTTCAGCTTA	GTTCATGCGA	AGAGCTTTTA	GTCTTCTCTT	77040
TAACTCTCTTA	GTCTTACGAA	TAGGAGAGAT	AGGCTCTCTT	TAGGCTTTCT	TGCTTCTCTT	77100
TTCGCTCTTA	TGTTGTAAGT	GACTCTCTTA	TGTATGAGAA	GTTCATGAGT	TGAGGCTTAT	77160
TGTATCTCTT	TAGCTCTCTT	AGCTGAGCTA	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77220
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77280
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77340
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77400
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77460
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77520
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77580
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77640
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77700
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77760
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77820
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77880
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	77940
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78000
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78060
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78120
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78180
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78240
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78300
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78360
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78420
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78480
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78540
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78600
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78660
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78720
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78780
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78840
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78900
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	78960
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79020
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79080
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79140
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79200
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79260
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79320
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79380
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79440
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79500
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79560
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79620
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79680
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79740
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79800
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79860
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79920
GTTCATCTTT	GTTCATCTTT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	TGTTGTAAGT	79980

The sequence shown in Figure 5S is a DNA sequence. The sequence is presented in a 10-column format. The first column contains the sequence of the first strand, and the second column contains the sequence of the second strand. The third column contains the sequence of the third strand, and the fourth column contains the sequence of the fourth strand. The fifth column contains the sequence of the fifth strand, and the sixth column contains the sequence of the sixth strand. The seventh column contains the sequence of the seventh strand, and the eighth column contains the sequence of the eighth strand. The ninth column contains the sequence of the ninth strand, and the tenth column contains the sequence of the tenth strand.

FIGURE 5T

[illegible]

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

2. The second step is to gather relevant information and data. This can involve research, consultation with experts, or collecting data from various sources.

3. The third step is to analyze the information and data collected. This involves identifying patterns, trends, and relationships that can help in understanding the problem.

4. The fourth step is to develop a solution or answer. This involves applying the knowledge and skills gained from the previous steps to create a response that addresses the problem.

5. The fifth step is to evaluate the solution or answer. This involves checking the work for accuracy, completeness, and relevance to the problem.

6. The sixth step is to communicate the solution or answer. This involves presenting the findings in a clear and concise manner, using appropriate language and format.

7. The seventh step is to reflect on the process. This involves thinking about what was learned from the experience and how it can be applied to future problems.

8. The eighth step is to seek feedback. This involves asking others for their thoughts and suggestions on the work, which can help in improving the quality of the solution.

9. The ninth step is to revise the solution or answer. This involves making changes to the work based on the feedback received, ensuring that it is the best possible response to the problem.

10. The tenth step is to submit the final solution or answer. This involves presenting the completed work to the appropriate authority or audience.

FIGURE 5U

TAGAGAGATT	TTGTGGGCTA	AAGCTGTCTT	TACCAAGAGA	TTTGGGTCTT	ATGTACACTT	34060
AAAGGATATA	AGGAGAGAAA	ATGAGAGAAA	TACAGCTCTG	ATGGATTTCT	TTTGGGGAATG	34120
AATCTAGAAA	ATTAAGAAAT	GATTATAAAT	AATGAAATAGG	GAAGCTGTAA	TAGAAATGACT	34180
TAGAGGCTTA	TTGAGAGGCT	TTAAAGATAT	ATGCTGTAAT	AAAGATAGTA	TTAGAAATATA	34240
TTAGAGAGAT	AAATAAATGC	ATTAAGATTA	AACATATGAA	ATAACATAT	ATCTAAGAAA	34300
AAAAAACTTA	GGAGTTTCTA	TAGAGAGATC	GGAGAGGATC	TTTCTGCTAT	AGTAGGGGAT	34360
TAGCTATATA	TTCTTAAAT	GGAAAACTGT	AGAAAAAAGG	AGCTTATGTA	TTTAAAGATA	34420
AGTTTCTCTG	AACTGTTTTT	TGATATCTCT	TGAAAGCTAT	TTTAAAGAT	ATATATCTAA	34480
TAACATAGAT	TTGATCTTCA	GAATTTGAGG	TGATTTCTCT	TTGTTGAAAT	TGAAATGAAA	34540
TTAAATAGAT	AGCTTTCTCT	AAAAATGGTG	TGATCTCTCT	TAAATAAAT	TTCTTTCTAT	34600
TAGCTATATA	TGATATCTTA	ATTTCTCTAT	ATTTCTCTAT	GAATGGATAC	ATTTGTCATG	34660
ATTAATAGGG	TTTCTTCTTA	AGGGCTGAGA	GGGTATTTCT	TTCTAACTGG	AAGCTGTGAC	34720
AGCTTCTGAG	AGGGCAATGG	AATCTTTCTT	AAATTTCTCT	AGCAAGAGAG	GGAGACACAG	34780
AAAGATCTGG	TTTCTACACT	TTTCTGGGAAC	TGAGAGAGGT	GAGAAAAAGG	GGATGTAAGT	34840
GGTCTCTCTA	TACAGAGAGG	AGGGTGGCTG	GGGTCTCTGG	TGGGGCTCTG	TTTCTCTAGAG	34900
AAATCTCTGG	AAAGGGATCT	GGGGTCTCTT	GGTTTGGACA	CAATTAAGCG	AAATAGGGGA	34960
AGGGTTCTCT	ATGGGGTCTT	TTTGGGGCTT	TTTGGACAGG	TTTGGACAGG	AGAGAGAGGA	35020
TGGGCAAGGA	TTGAAGAGGG	AGTGGGTGAG	AGGTTTCTTA	GGTCTATGGA	AGTTCTGAA	35080
TTGGATCTGA	AGAGAGAT	ATTAATCTCT	TGCAATGTAAC	AGCTAGAGAT	AGCTGGGAAT	35140
TGAGAGGGA	TAAAGATCTG	TTTCTCTCTT	AGTGGCTCT	AGCTGGCTCT	GACTATGGAG	35200
AGGTCTTGAA	TTGTAGCTCT	CAAGTGGGGG	AGAAATAGGC	TGGGAGAGAG	AGGGGAGAGG	35260
AGGAACTACA	GGGTCTCTCT	GGAGCTCTTA	AGAGAAAGGA	GTAAAAATG	AGGGATGGAG	35320
AAAAATATTA	GGCTAAATTA	AGTTTCTGTA	TTTCTCTGAT	ATGAGGCTG	AGCTGACAGT	35380
GAAGATCTTA	TTGCTTCTTA	ATGCTCTCTT	TTGAGAGGCT	GGATGAGGCT	GGATGAGGCT	35440
GAAGATCTTA	TTGCTTCTTA	TTGCTTCTTA	TTGCTTCTTA	TTGCTTCTTA	TTGCTTCTTA	35500
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	35560
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	35620
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	35680
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	35740
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	35800
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	35860
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	35920
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	35980
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36040
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36100
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36160
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36220
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36280
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36340
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36400
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36460
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36520
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36580
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36640
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36700
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36760
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36820
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36880
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	36940
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	37000
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	37060
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	37120
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	37180
TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	TTTCTCTTA	37240
TT						

FIGURE 6

GTATAAAGTT AGTAAATGTS AGGCCTCTCT CGATGCCTGG GTCTTGGGCT TTGGTTCTCA	60
GTCTCTGNTA AATCATCTCT CTGGAGGAGA AGACCTTAG ATCTGGCTCT TCTCAGGGGC	120
ATTTTAAAGA CAAATGAAAA TAAA ATG GAA ACC ACT TCA CTA CAG CGG AAA	171
Met Glu Thr Thr Ser Leu Gln Arg Lys	
1 5	
TTT CCA GAA TGG ATG TCT ATG CAG AGT CAA AGA TGT GCT ACA GAA GAA	219
Phe Pro Glu Trp Met Ser Met Gln Ser Gln Arg Cys Ala Thr Glu Glu	
10 15 20 25	
AAG GCG TGG GTT CAG AAG AGT GTT CTT GAA GAT AAG CTC GGA TTC TTA	267
Lys Ala Cys Val Gln Lys Ser Val Leu Glu Asp Asn Leu Pro Phe Leu	
30 35 40	
GAA TTC CTT GGA TCC ATT GTT TAC AGT TAT GAA CTT AGT GAT TGC TCC	315
Glu Phe Pro Gly Ser Ile Val Tyr Ser Tyr Glu Ala Ser Asp Cys Ser	
45 50 55	
TTC CTC TCT GAA GAC ATT AGC ATG CTT CTC TCT GAT GGC GAT CTC GTC	363
Phe Leu Ser Glu Asp Ile Ser Met Arg Leu Ser Asp Gly Asp Val Val	
60 65 70	
GGA TTT GAC ATG GAA TGG GCG GCG ATA TAC AAG CCA GGG AAA AGA AGC	411
Gly Phe Asp Met Glu Trp Pro Pro Ile Tyr Lys Pro Gly Lys Arg Ser	
75 80 85	
AGA CTC GCA CTC ATT CAG TTS TGT CTG TCT GAG AGC AAA TGT TAC TTG	459
Arg Val Ala Val Ile Gln Leu Cys Val Ser Glu Ser Lys Cys Tyr Leu	
90 95 100 105	
TTT CAG ATT TCT TCC ATG TCA GTT TTC CCG CAG GGA TTA AAA ATG TTA	507
Phe His Ile Ser Ser Met Ser Val Phe Pro Gln Gly Leu Lys Met Leu	
110 115 120	
CTA GAA AAG AAA TCA ATT AAG AAG GCA GGG GTT GGG ATT GAA GGG GAC	555
Leu Glu Asn Lys Ser Ile Lys Lys Ala Gly Val Gly Ile Glu Gly Asp	
125 130 135	
CAG TTS AAA CTT CTC CTT GAT TTT GAC CTC AAG TTG GAG AGT TTT CTC	603
Gln Trp Lys Leu Leu Arg Asp Phe Asp Val Lys Leu Glu Ser Phe Val	
140 145 150	
GAG CTC AAG GAT GTT GCG AAT GAA AAG TTG AAG TCC GCA GAG ACC TGG	651
Glu Leu Thr Asp Val Ala Asn Glu Lys Leu Lys Cys Ala Glu Thr Trp	
155 160 165	
AGC CTC AAT GGT CTC GTT AAA CAG CTC TTA GCG AAA CAA TTT TTG AAA	699
Arg Ser Asn Gly Thr Val Lys CAG CTC TTA GCG AAA CAA TTT TTG AAA	
170 175 180	

FIGURE 6 (CONT.)

GAC CAG AAA CTG TAT GCA GCG ACT GAT GCT TAT GCT GGT GTT ATC ATC Asp Gln Lys Leu Tyr Ala Ala Thr Asp Ala Tyr Ala Gly Leu Ile Ile 205 210 215	795
TAT CAA AAA TTA GGA AAT TTG GGT GAT ACT GCG CAA GTG TTT GCT CTA Tyr Gln Lys Leu Gly Asn Leu Gly Asp Thr Ala Gln Val Phe Ala Leu 220 225 230	843
AAT AAA GCA GAG GAA AAC CTA CCT CTG GAG ATG AAG AAA CAG TTG AAT Asn Lys Ala Glu Glu Asn Leu Pro Leu Glu Met Lys Lys Gln Leu Asn 235 240 245	891
TCA ATC TCG GAA GAA ATG AGG GAC CTA GCG AAT CGT TTT CCT GTC ACT Ser Ile Ser Glu Glu Met Arg Asp Leu Ala Asn Arg Phe Pro Val Thr 250 255 260 265	939
TGC AGA AAT TTG GAA ACT CTC GAG AGG GTT CCT GTA ATA TTG AAG AGT Cys Arg Asn Leu Glu Thr Leu Gln Arg Val Pro Val Ile Leu Lys Ser 270 275 280	987
ATT TCA GAA AAT CTC TGT TCA TTG AGA AAA GTG ATC TGT GGT CTT ACA Ile Ser Glu Asn Leu Cys Ser Leu Arg Lys Val Ile Cys Gly Pro Thr 285 290 295	1035
AAC ACT GAG ACT AGA CTC ACG CCG GCG AGT AGT TTT AAT TTA CTC TCA Asn Thr Glu Thr Arg Leu Lys Pro Gly Ser Ser Phe Asn Leu Leu Ser 300 305 310	1083
TCA GAG GAT TCA GGT GGT GGT GGA GAA AAA GAG AAA CAG ATT GGA AAA Ser Glu Asp Ser Ala Ala Ala Gly Glu Lys Glu Lys Gln Ile Gly Lys 315 320 325	1131
CAT AGT ACT TTT GCT AAA ATT AAA GAA GAA CCA TGG GAC CCA GAA CTT His Ser Thr Phe Ala Lys Ile Lys Glu Glu Pro Trp Asp Pro Glu Leu 330 335 340 345	1179
GAC ACT TTA GTG AAG CAA GAG GAG GTT GAT GTA TTT AGA AAC CAA GTG Asp Ser Leu Val Lys Gln Glu Glu Val Asp Val Phe Arg Asn Gln Val 350 355 360	1227
AAG CAA AAA AAA GGT CAA TGT GAA AAT GAA ATA GAA GAC AAT CTG TTG Lys Gln Glu Lys Gly Glu Ser Glu Asn Glu Ile Glu Asp Asn Leu Leu 365 370 375	1275
AGA GAA GAT ATG GAA AGA ACT TGT GTG ATT CGT AGT ATT TCA GAA AAT Arg Glu Asp Met Glu Arg Thr Cys Val Ile Pro Ser Ile Ser Glu Asn 380 385 390	1323
GAA ATT CAA GAT TTG GAA TAA CAA GCT AAA GAA GAA AAA TAT CAT CAT Glu Ile CAA GAT TTG GAA TAA CAA GCT AAA GAA GAA AAA TAT CAT CAT 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995	1371

FIGURE 6 (CONT.)

GAC	TCC	TCC	CAT	ATA	ATT	GAA	AGT	GAT	GAA	GAT	TTG	GAA	ATG	GAG	ATG	1467
Asp	Ser	Ser	Tyr	Ile	Ile	Glu	Ser	Asp	Glu	Asp	Leu	Glu	Met	Glu	Met	
			430						435					440		
CTG	AAG	TCT	TTA	GAA	AAC	CTA	AAT	AGT	GAC	GTG	GTG	GAA	CCC	ACT	CAC	1515
Leu	Lys	Ser	Leu	Glu	Asn	Leu	Asn	Ser	Asp	Val	Val	Glu	Pro	Thr	His	
			445					450					455			
TCT	ACA	TGG	TTG	GAA	ATG	GGA	ACC	AAT	GGG	CGT	CTT	CCT	ECT	GAG	GAG	1563
Ser	Thr	Trp	Leu	Glu	Met	Gly	Thr	Asn	Gly	Arg	Leu	Pro	Pro	Glu	Glu	
		460					465					470				
GAA	GAT	GGA	CAC	GGA	AAT	GAA	GCC	ATC	AAA	GAG	GAG	CAG	GAA	GAA	GAG	1611
Glu	Asp	Gly	His	Glv	Asn	Glu	Ala	Ile	Lys	Glu	Glu	Gln	Glu	Glu	Glu	
	475					480						485				
GAC	CAT	TTA	TTG	CCC	GAA	CCC	AAC	GCA	AAG	GAA	ATT	AAT	TGC	CTC	AAG	1659
Asp	His	Leu	Leu	Pro	Glu	Pro	Asn	Ala	Lys	Gln	Ile	Asn	Cys	Leu	Lys	
490				495					500				505			
ACC	TAT	TTG	GGA	CAC	AGC	ACT	TTT	AAA	CCC	GTT	CAG	TGC	AAA	CTC	ATC	1707
Thr	Tyr	Phe	Gly	His	Ser	Ser	Phe	Lys	Pro	Val	Gln	Trp	Lys	Val	Ile	
			510					515					520			
CAT	TCT	GTA	TTA	GAA	GAG	AGA	AGA	GAT	AAT	GTT	GTT	GTC	ATG	GCA	ACT	1755
His	Ser	Val	Leu	Glu	Glu	Arg	Arg	Asp	Asn	Val	Val	Val	Met	Ala	Thr	
			525					530					535			
GGA	TAT	GGG	AAG	ACT	CTG	TGC	TTG	CAG	TAT	CCG	CGT	GTT	TAT	ACA	GGC	1803
Gly	Thr	Gly	Lys	Ser	Leu	Cys	Phe	Gln	Tyr	Pro	Phe	Val	Tyr	Thr	Gly	
		540				545						550				
AAG	ATT	GGC	ATT	GTC	ATT	TCA	CGT	CTC	ATT	TCC	TTA	ATG	GAA	GAC	CAA	1851
Lys	Ile	Gly	Ile	Val	Ile	Ser	Pro	Leu	Ile	Ser			Glu	Asp	Gln	
	555					560				565						
GTC	CTC	GGG	TTT	GAG	CTG	TCC	AAT	GTT	GCA	GGC	TCT	TTA	GTT	GGA	TCT	1899
Val	Leu	Gln	Leu	Glu	Leu	Ser	Asn	Val	Pro	Ala	Cys	Leu	Leu	Gly	Ser	
570				575					580					585		
GCA	CAG	TCA	AAA	AAT	ATT	CTA	GGA	GAT	CTT	AAA	TTA	GGC	AAA	TAT	AGG	1947
Ala	Gln	Ser	Lys	Asn	Ile	Leu	Gly	Asp	Val	Lys	Leu	Gly	Lys	Tyr	Arg	
			590					595					600			
GTC	ATC	TAC	ATA	ACT	GCA	GAG	TTC	TGT	TCT	GGT	AAC	TTG	GAT	CTA	CTC	1995
Val	Ile	Tyr	Ile	Thr	Pro	Glu	Phe	Cys	Ser	Gly	Asn	Leu	Asp	Leu	Leu	
			605					610					615			
GAG	GAA	TTT	GAG	TCT	ACT	AAT	CTC	ATC	ACT	CTC	ATT	CTT	GTC	GAT	GAA	
620																

FIGURE 6 (CONT.)

ATG CTG GGC TGT GTT AAA ACA GCG CTC CCA TTG GTT CCA GTC ATT GCA	2139
Met Leu Gly Ser Leu Lys Thr Ala Leu Pro Leu Val Pro Val Ile Ala	
650 655 660 665	
CTC TCC GGT ACT GCA AGC TCT TCC ATC CGG GAA GAC ATT ATA AGC TGC	2187
Leu Ser Ala Thr Ala Ser Ser Ser Ile Arg Glu Asp Ile Ile Ser Cys	
670 675 680	
TTA AAC CTG AAA GAC GGT CAG ATC ACC TGC ACT GGA TTT GAT CGG CCA	2235
Leu Asn Leu Lys Asp Pro Gln Ile Thr Cys Thr Gly Phe Asp Arg Pro	
685 690 695	
AAT CTG TAC TTA GAA GTT GGA CGG AAA ACA GGG AAC ATC CTT CAG GAT	2283
Asn Leu Tyr Leu Glu Val Gly Arg Lys Thr Gly Asn Ile Leu Gln Asp	
700 705 710	
ATA AAG CCG TTT GTG CTC CGA AAG GCA AGT TGT GGC TGG GAA TTT GAA	2331
Leu Lys Pro Phe Leu Val Arg Lys Ala Ser Ser Ala Trp Ile Phe Glu	
715 720 725	
GST CCA AGC ATC ATC TAT TGT GGT TCG AGA AAA ATG ACA GAA CAA GTT	2379
Gly Pro Thr Ile Ile Tyr Cys Pro Ser Arg Lys Met Thr Gln Gln Val	
730 735 740 745	
ACT GGT GAA GTT GGG AAA CTG AAC TTA GCG TGG AGA ACA TAC CAC GGT	2427
Thr Ala Glu Leu Gly Lys Leu Asn Leu Ala Cys Arg Thr Tyr His Ala	
750 755 760	
GGC ATG AAA ATT ACC GAA AGG AAG GAC GTT CAT CAT AGG TTC CTG AGA	2475
Gly Met Lys Ile Ser Glu Arg Lys Asp Val His His Arg Phe Leu Arg	
765 770 775	
GAT GAA ATT CAG TGT GTT GTA GGT ACT GTA GGT TTT GGA ATG GGC ATT	2523
Asp Glu Ile Gln Cys Val Val Ala Thr Val Ala Phe Gly Met Gly Ile	
780 785 790	
AAT AAA GGT GAC ATT TGG AAA GTT ATT CAT TAT GGT GCG GGT AAG GAA	2571
Asn Lys Ala Asp Ile Arg Lys Val Ile His Tyr Gly Ala Pro Lys Glu	
795 800 805	
ATG GAA TCC TAT TAC CAG GAA ATT GST AGA GGT GGC CGG GAT GGA GTT	2619
Met Glu Ser Tyr Tyr Gln Glu Ile Gly Arg Ala Gly Arg Asp Gly Leu	
810 815 820 825	
CAG AGT TCC TGT CAC TTG CTC TGG GGT CCA GCA GAC TTT AAC AAG TCC	2667
Gln Ser Ser Cys His Leu Leu Trp Ala Pro Ala Asp Phe Asn Thr Ser	
830 835 840	
AGG AAT CTC GTT ATT GAG ATT CAG GAT GAA AAG TTC TGG TTA TAT AAA	2715

FIGURE 6 (CONT.)

AGG CGA CGA ATC ATC TTG TCC CAT TTT GAG GAC AAA TGT CTG CAG AAG	2811
Arg Arg Arg Ile Ile Leu Ser His Phe Glu Asp Lys Cys Leu Gln Lys	
375 880 885	
GCC TCC TTG GAC ATT ATG GGA ACT GAA AAA TGC TGT GAT AAT TCC AGG	2859
Ala Ser Leu Asp Ile Met Gly Thr Glu Lys Cys Cys Asp Asn Cys Arg	
890 895 900 905	
CCC AGG CTG AAT CAT TGC ATT ACT GCT AAC AAC TCA GAG GAC GCA TCC	2907
Pro Arg Leu Asn His Cys Ile Thr Ala Asn Asn Ser Glu Asp Ala Ser	
910 915 920	
CAA GAC TTT GGG CCA CAA GCA TTC CAG CTA CTG TCT GGT GTG GAC ATC	2955
Gln Asp Phe Gly Pro Gln Ala Phe Gln Leu Leu Ser Ala Val Asp Ile	
925 930 935	
CTG CAG GAG AAA TTT GGA ATT GGG ATT CCG ATC TTA TTT CTC CCA GGA	3003
Leu Gln Gln Lys Phe Glu Ile Gly Ile Pro Ile Leu Phe Leu Arg Gly	
940 945 950	
TCT AAT TCT CAG CTT CTT CTT GAT AAA TAT CCG GGT CAC AGG CTC TTT	3051
Ser Asn Ser Gln Arg Leu Pro Asp Lys Tyr Arg Gly His Arg Leu Phe	
955 960 965	
GCT GCT GCA AAG CAG CAA GTA GAA AGT TCG TCG AAG ACC CTT TCT CAC	3099
Gly Ala Gly Cys Gln Gln Ala Glu Ser Trp Trp Lys Thr Leu Ser His	
970 975 980 985	
CAT CTC ATA GCT GAA GGA TTC TTG GTA GAA GTT CCG AAG GCA AAC AAA	3147
His Leu Ile Ala Gln Gly Phe Leu Val Glu Val Pro Lys Glu Asn Lys	
990 995 1000	
TAT ATA AAG ACA TGT TCC CTC ACA AAA AAG GGT AGA AAG TGG CTT GGA	3195
Tyr Ile Lys Thr Cys Ser Leu Thr Lys Lys Gly Arg Lys Trp Leu Gly	
1005 1010 1015	
GAA GGT AGT TCG CAG TCT CTT CCG AGC CTT CTC CTT CAA GTT AAT GAA	3243
Glu Ala Ser Ser Gln Ser Pro Pro Ser Leu Leu Leu Gln Ala Asn Glu	
1020 1025 1030	
GAG ATG TTT TCA AGG AAA GTT CTG CTA CCA AGT TCT AAT GGT GTA TGT	3291
Glu Met Phe Pro Arg Lys Val Leu Leu Pro Ser Ser Asn Pro Val Ser	
1035 1040 1045	
CCA GAA AGG AGG CAA CAT TCC TCT AAT CAA AAC CCA GGT GGA TTA ACT	3339
Pro Glu Thr Thr Gln His Ser Ser Asn Gln Asn Pro Ala Gly Leu Thr	
1050 1055 1060 1065	
ACC AAG CAG TTT AAT TTG GAG AGA AGC CAT TTT TAC AAA GTG CTT CAG	3387
Thr Lys Gln Phe Asn Cys GAG AGA AGC CAT TTT TAC AAA GTG CTT CAG	

FIGURE 6 (CONT.)

TCA CCA GGA ACA TCT TCC AGC CCC TTA GAA CCT GCC ATC TCA GCC CAA Ser Pro Gly Thr Ser Ser Ser Pro Leu Glu Pro Ala Ile Ser Ala Gln 1100 1105 1110	3483
GAG CTG GAC GGT CCG ACT GGG CTA TAT GCC AGG CTG GTG GAA GCA AGG Glu Leu Asp Ala Arg Thr Gly Leu Tyr Ala Arg Leu Val Glu Ala Arg 1115 1120 1125	3531
CAG AAA CAC GGT AAT AAG ATG GAT GTA CCT CCA GCT ATT TTA GCA ACA Gln Lys His Ala Asn Lys Met Asp Val Pro Pro Ala Ile Leu Ala Thr 1130 1135 1140 1145	3579
AAC AAG GTT CTG CTG GAC ATG GCT AAA ATG AGA CCC ACT ACT GTT GAA Asn Lys Val Leu Leu Asp Met Ala Lys Met Arg Pro Thr Thr Val Glu 1150 1155 1160	3627
AAC ATG AAA CAG ATC GAC GGT GTG TCT GAA GCC AAA GGT GGT CTG TTG Asn Met Lys Gln Ile Asp Gly Val Ser Glu Gly Lys Ala Ala Leu Leu 1165 1170 1175	3675
GCC GGT CTG TTG GAA GTG ATC AAA GAT TTC TGT CAA GTA ACT ACT GTT Ala Pro Leu Leu Glu Val Ile Lys His Phe Cys Gln Val Thr Ser Val 1180 1185 1190	3723
CAG ACA GAC CTC GTT TCC ACT GCC AAA GGT CAC AAG GAA CAG GAG AAA Gln Thr Asp Leu Leu Ser Ser Ala Lys Pro His Lys Gln Gln Glu Lys 1195 1200 1205	3771
AGT CAG CAG ATG GAA AAG AAA GAC TGC TCA CTC GCC CAG TCT GTG GCC Ser Gln Glu Met Glu Lys Lys Asp Cys Ser Leu Pro Gln Ser Val Ala 1210 1215 1220 1225	3819
CTC ACA TAC ACT CTA TTC CAG GAA AAG AAA ATG GCC TTA CAC AGC ATA Val Thr Tyr Thr Leu Phe Gln Glu Lys Lys Met Pro Leu His Ser Ile 1230 1235 1240	3867
GCT CAG AAC AGG CTC CTG GGT CTC ACA GCA GCC GGC ATG CAG TTA GCC Ala Glu Asn Arg Leu Leu Pro Leu Thr Ala Ala Gly Met His Leu Ala 1245 1250 1255	3915
CAG CCG CTG AAA GCC GCG TAC CCC CTG GAT ATG GAG CGA GGT GGC CTG Gln Ala Val Lys Ala Gly Tyr Pro Leu Asp Met Glu Arg Ala Gly Leu 1260 1265 1270	3963
ACC CCA GAG ACT TGG AAG ATT ATT ATG GAT GTC ATC CGA AAC GGT CCC Thr Pro Glu Thr Trp Lys Ile Ile Met Asp Val Ile Arg Asn Pro Pro 1275 1280 1285	4011
ATC AAC TCA GAT ATG TAT AAA GTT AAA CTC ATC AGA ATG TTA GTT GGT Ile Asn Thr GAT ATG TAT AAA GTT AAA CTC ATC AGA ATG TTA GTT GGT 1290 1295 1300 1305	4059

FIGURE 6 (CONT.)

AGT GGT TCC GAC ACC AGA ACC CAG CCT CCT TGT GAT TCC AGC AGG AAG Ser Gly Ser Asp Ser Arg Thr Gln Pro Pro Cys Asp Ser Ser Arg Lys 1325 1330 1335	4155
AGG CGT TTC CCC AGC TCT GCA GAG AGT TGT GAG AGC TGT AAG GAG AGC Arg Arg Phe Pro Ser Ser Ala Glu Ser Cys Glu Ser Cys Lys Glu Ser 1340 1345 1350	4203
AAA GAG GCG GTC ACC GAG ACC AAG GCA TCA TCT TCA GAG TCA AAG AGA Lys Glu Ala Val Thr Glu Thr Lys Ala Ser Ser Ser Glu Ser Lys Arg 1355 1360 1365	4251
AAA TTA CCC GAG TGG TTT GCC AAA GGA AAT GTC CCC TCA GCT GAT ACC Lys Leu Pro Glu Trp Phe Ala Lys Gly Asn Val Pro Ser Ala Asp Thr 1370 1375 1380 1385	4299
GGC AGC TCA TCA TCA ATG GGC AAG ACC AAA AAG AAA GGT CTC TTT AGT Gly Ser Ser Ser Ser Met Ala Lys Thr Lys Lys Lys Gly Leu Phe Ser 1390 1395 1400	4347
TAAATGACG ACAGTGGAGC AATTTGTGTG TCGTACATCT TCATTGCTAT AAAGAATGAA 4407	
NAGAAATATT TTAAGCTCAA AATTATTCAA AGTCCAAAGT GAAGCTCACC TAAAGCTCGA 4467	
GCCATAGAGT CTTTAAATTG CCGTTGCGAG TTGAGCTACA GTATCTGAAC CTTCTGAGAC 4527	
CCGGAGTGA SCATAGACTG TGAAGTCGGC TCGCTTTCCG ATTGCGCTCC GAACCCGTGT 4587	
CAGTGTCAAG TTGAGTCTT TCTCTCTCTG CAGCAGTGTG TCTTGGAAAT GGAGGCTGTG 4647	
TGCTTTTGAC ATATAGAACA GATCAATAT TGCATAGGGA CAGATATGAA GATNCAGCG 4707	
GTCTTTGCTT TCTTATGAG ATGCGTCTAT GACATATCA GTGCACCAGC CCGGCCAGGG 4767	
AGACATCAGC TTGCATTTAA AAAGG 4792	

FIGURE 7

Genomic sequence

001459 001459

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

1. The first step is to identify the problem. This involves understanding the current situation and the goals that need to be achieved.

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840. 84

the 1990s, the number of people in the world who are under 15 years of age is expected to increase from 1.1 billion to 1.5 billion. The number of people aged 65 and over is expected to increase from 200 million to 400 million. The number of people aged 15 and over is expected to increase from 3.5 billion to 4.5 billion. The number of people aged 15 and over is expected to increase from 3.5 billion to 4.5 billion. The number of people aged 15 and over is expected to increase from 3.5 billion to 4.5 billion.

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

FIGURE 7 (CONT.)

[illegible]

FIGURE 7 (CONT.)

[illegible]

[illegible]

[illegible]

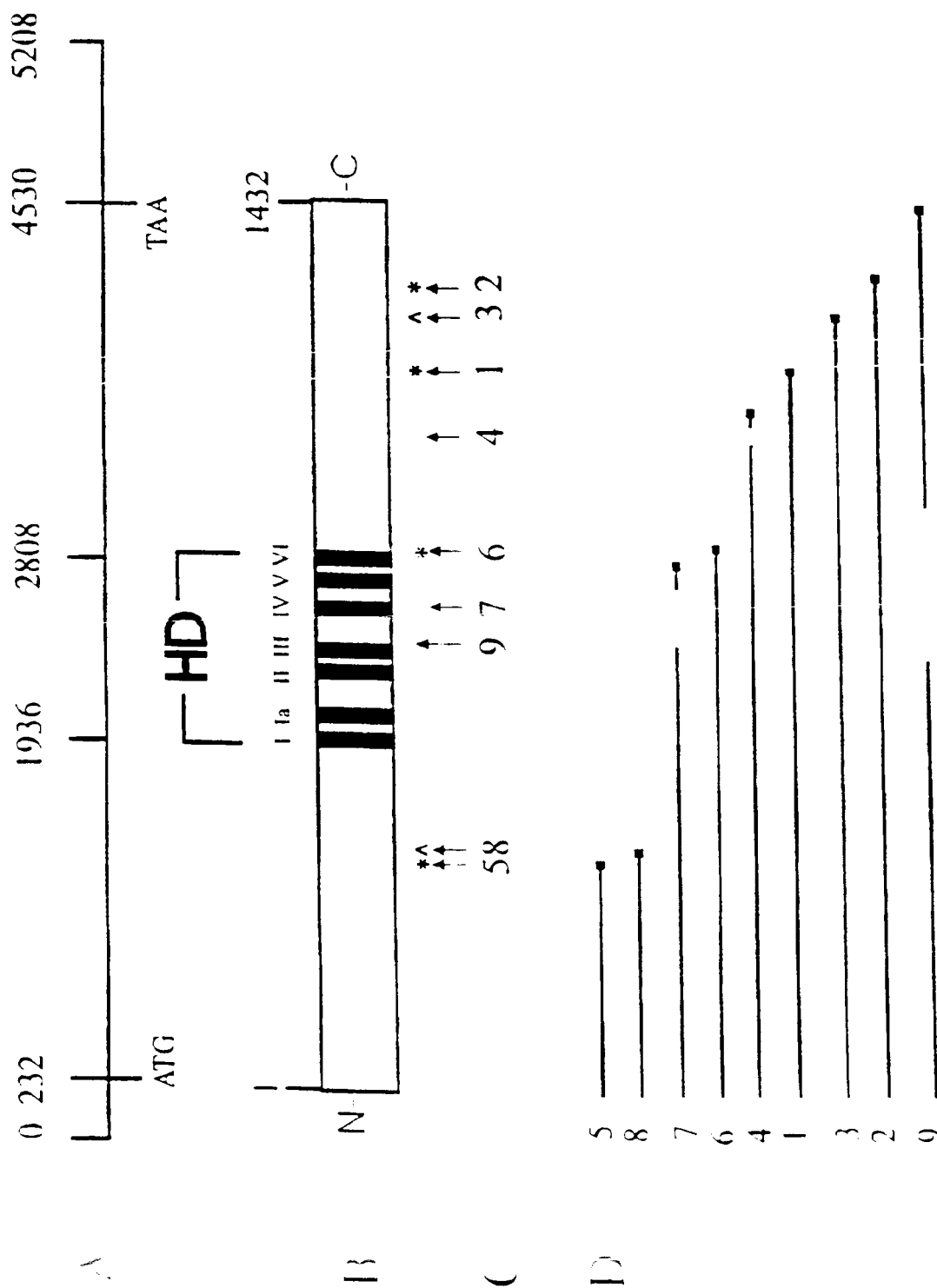


FIGURE 8

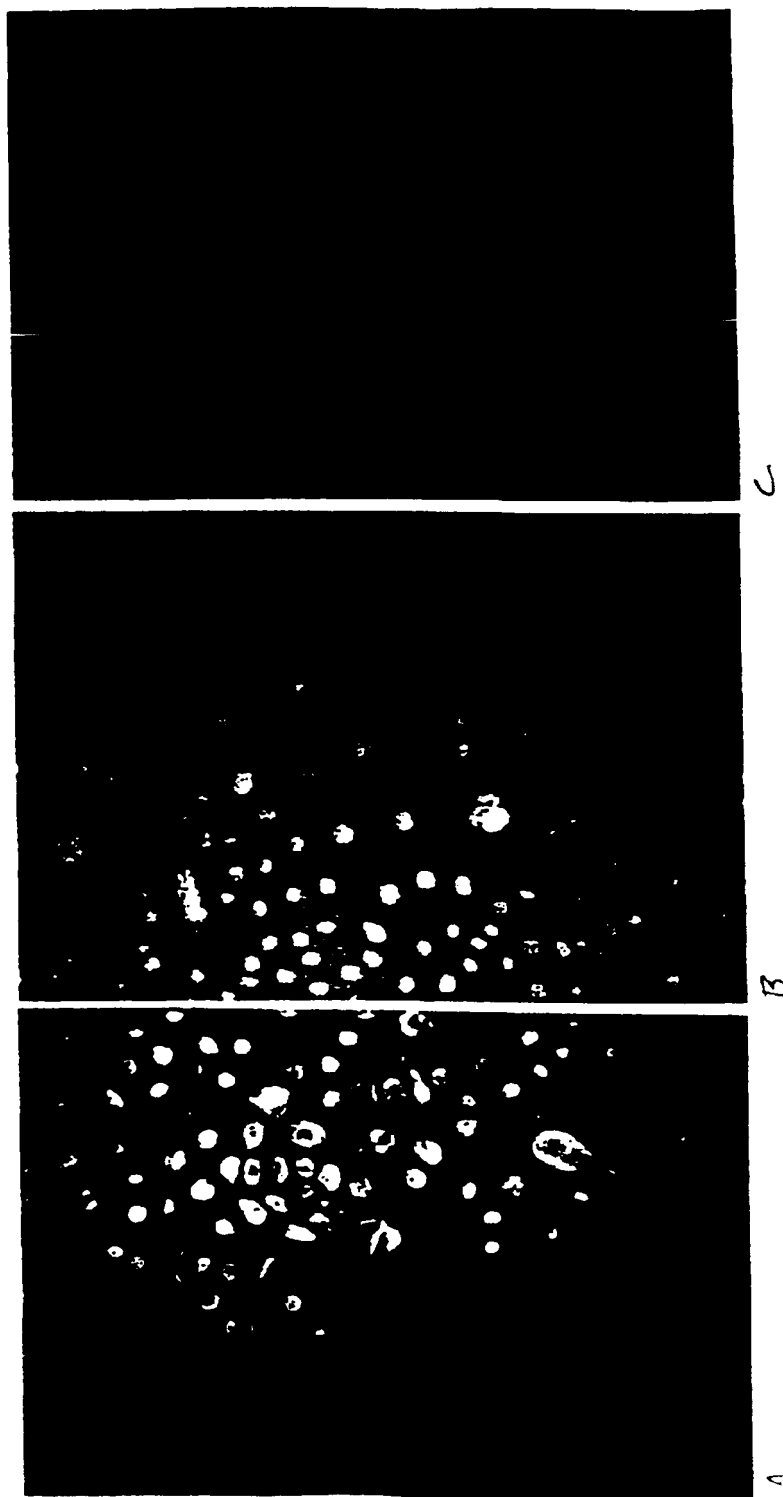


Fig. 6

Effect of the WRN protein by indirect immunofluorescence in adherent human epithelial cells. Cells from prostate epithelium were placed on a glass slide and allowed to grow for several days before staining. The cells were fixed, permeabilized, stained simultaneously for the WRN protein, DNA in the nucleus, or F-actin in the cytoplasm. Panel A shows cells after staining with an appropriate dilution of the anti-WRN p8-pink (rabbit antiserum). The secondary antibody (goat anti-mouse IgG) was labeled with FITC. Panel B shows cells in the next culture vessel after staining nuclear DNA with bis benzimidazole Panel C shows cells after staining with BODIPY-tagged phalloidin. 200X magnification.

FIGURES 9A, B and C

FIGURE 10

Human-Mouse Annotated Protein Alignment

| | | EXONS 10 & 11 | |
|-------|----|--------------------------|-----------------------|
| | | Human | Mouse |
| | | Download seq | files: Human or Mouse |
| Human | 10 | IS LQRI FEMMS MUSQRCATKE | 100 |
| Human | 11 | TA CQKQCEWHN VQNCACVLE | 100 |
| Human | 12 | AS IK SFSEIS MSLSRHVVG | 100 |
| Human | 13 | AS IK SFSEIS MSLSRHVVG | 100 |
| Human | 14 | Y LPHVSSKSVF PQGKHILN | 100 |
| Human | 15 | Y LPHVSSKSVF PQGKHILN | 100 |
| Human | 16 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 17 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 18 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 19 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 20 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 21 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 22 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 23 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 24 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 25 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 26 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 27 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 28 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 29 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 30 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 31 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 32 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 33 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 34 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 35 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 36 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 37 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 38 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 39 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 40 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 41 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 42 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 43 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 44 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 45 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 46 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 47 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 48 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 49 | E LPHVSSKSVF PQGKHILN | 100 |
| Human | 50 | E LPHVSSKSVF PQGKHILN | 100 |

1050 EBYSDIANK STEHLER/1975/11/15/1506/1506

[illegible]

FIGURE 10
(CONTINUED)

| | | |
|-------|---|------|
| Human | QVDF QVQAFKILLSA VVLIAGEFGI QLPFLFLLNS HSQRKLAQYR | 1050 |
| Human | AKRS QAESWMTLS IMLIAEGFLV EVFEEHYIK TQSLTEGRK | 1050 |
| Human | TURD QTESWKAFS RQIITEGFLV EVSRVDFHE TQALTERHN | |
| Human | QSP FQILIQANEE MFRKVLVLS ERFVDFHTQD IRRKQHPAGL | 1100 |
| Human | RESL QULIQANEE LAFKRFLLPS ERFVDFHTQD IRRKQHPAGL | |
| Human | QER TUSYKVEFLV SSGNIPEKS AGRHSQFSS SLEPALSQ | 1150 |
| Human | QER LYSKPCIKI SSGNISKS THUSPEAY SSSQVISAQ | |
| Human | QYA NIVEARQKIA NENQUPPATL ATKVLLVIA KRRPTVFM | 1200 |
| Human | QYD KIVEARQKIA NENQUPPATL ATKVLLVIA KRRPTVFM | |
| Human | QGR AALAPLELV IRIFCQTVS QTHLSSAKP IREKERSQFM | 1250 |
| Human | QGR AALAPLELV IRIFCQTVS QTHLSSAKP IREKERSQFM | |
| Human | QUS VAVTYTIFUE KRPILISIAE QHLLPTAAG IMLAQAVKAI | 1300 |
| Human | QUS HAITYSIFUE KRPILISIAE QHLLPTAAG IMLAQAVKAI | |
| Human | QAL TPETKILMD VIRNPINSQ HVEKLLRMV VPRHLPYLI | 1350 |
| Human | QAL TPETKILMD VIRNPINSQ HVEKLLRMV VPRHLPYLI | |
| Human | QRI DISRTUPQD SSRKHFFSS AESSESSES KEAVTETRA | 1400 |
| Human | QRI DISRTUPQD SSRKHFFSS AESSESSES KEAVTETRA | |
| Human | QLP ENFAGNUPS ADTGSSSSMA KTKKGLFS* | 1440 |
| Human | QLP ENFAGNUPS ADTGSSSSMA KTKKGLFS* | |

Analysis and sequence generated by Δ.SWIZ

Last Update 08-03-96

INTERNATIONAL SEARCH REPORT

International Application No.

PLT/US 96/20785

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/00 C12N15/61 C12N9/90 C12N15/85 C12N15/86
C12N5/10 C12N5/20 C07K16/40 C12Q1/68 A01K67/027
G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K C12Q A01K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|-----------------------|
| X | EMBL SEQUENCE DATABASE,
30 January 1995, HEIDELBERG, BRD,
XP002032088
F. BOUILLAUD: "Study of expressed
sequence tags in adipose tissue 1995"
accession no. T39125
see abstract
--- | 2,15,16 |
| X | EMBL SEQUENCE DATABASE,
29 May 1995, HEIDELBERG, BRD, XP002032089
R. BERRY ET AL.: "Gene-based sequence
tagged sites (STSs) as the basis for a
human gene map"
see accession no. R58879
see abstract

-/- | 2,15,16 |

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"I" document which may throw doubts on priority claims) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure (see, e.g., Article 15 of the EPC)

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more prior art documents

Signature of the International Searching Authority

Signature of the International Searching Authority

Date of completion of the international search

Date of mailing of the international search report

2 June 1997

10.06.97

Name and mailing address of the ISA

European Patent Office, P.O. Box 1, 7000 Lausanne, Switzerland
Tel. (41) 21 611 91 11
Fax (41) 21 611 91 99

Authorized officer

Authorized officer

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 96/20785

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|-----------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | <p>NATURE,
vol. 355, 20 February 1992, MACMILLAN
JOURNALS LTD., LONDON, UK,
pages 735-738, XP002032090
M. GOTO ET AL.: "Genetic linkage of
Werner's syndrome to five markers on
chromosome 8"
cited in the application
see the whole document</p> <p>---</p> | 1-21 |
| A | <p>AMERICAN JOURNAL OF HUMAN GENETICS 55 (2).
1994. 356-364. ISSN: 0002-9297,
August 1994, XP000674409
YU C-E ET AL: "Linkage disequilibrium and
haplotype studies of chromosome 8p
11.1-2.1 markers and Werner syndrome."
cited in the application
see the whole document</p> <p>---</p> | 1-21 |
| A | <p>ANNUAL MEETING OF THE AMERICAN SOCIETY FOR
BIOCHEMISTRY AND MOLECULAR BIOLOGY, SAN
FRANCISCO, CALIFORNIA, USA, MAY 21-25,
1995. FASEB JOURNAL 9 (6). 1995. A1270.
ISSN: 0892-6638,
24 April 1995, XP002032091
THWEATT R ET AL: "A novel cDNA
overexpressed in Werner syndrome (WS)
fibroblasts inhibits colony formation in
normal human fibroblasts and in HeLa
cells."
abstract no.84
see abstract</p> <p>---</p> | 1-21 |
| A | <p>GENOMICS 28 (2). 1995. 147-153. ISSN:
0888-7543,
20 July 1995, XP000674353
KURIMASA A ET AL: "Construction of 110
cosmid markers and a 4.5-Mb YAC contig on
human chromosome 8p12-q11."
see the whole document</p> <p>---</p> | 1-21 |
| A | <p>GENOMICS 28 (3). 1995. 566-569. ISSN:
0888-7543,
10 August 1995, XP000674354
YE L ET AL: "Genetic association between
chromosome 8 microsatellite (MS8-134) and
Werner Syndrome (WRN): Chromosome
microsatellite linkage and homozygosity mapping of</p> | 1-21 |

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 96/20785

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|----------|--|-----------------------|
| A | 45TH ANNUAL MEETING OF THE AMERICAN SOCIETY OF HUMAN GENETICS, MINNEAPOLIS, MINNESOTA, USA, OCTOBER 24-28, 1995. AMERICAN JOURNAL OF HUMAN GENETICS 57 (4 SUPPL.). 1995. A266. ISSN: 0002-9297, October 1995, XP000674410
NAKURA J ET AL: "Genetic association between chromosome 8 microsatellites and Werner syndrome (WRN)."
abstract no. 1544
see abstract
--- | 1-21 |
| A | NUCLEIC ACIDS RESEARCH,
vol. 22, no. 22, 11 November 1994, IRL PRESS LIMITED, OXFORD, ENGLAND,
pages 4566-4573, XP002032092
M. SEKI ET AL.: "Molecular cloning of cDNA encoding human DNA helicase Q1 which has homology to Escherichia coli Rec Q helicase and localization of the gene at chromosome 12p12"
see the whole document
--- | 1-21 |
| A | J. BIOL. CHEM.,
vol. 269, no. 47, 25 November 1994, AM. SOC. BIOCHEM. MOL. BIOL., INC., BALTIMORE, US,
pages 29838-29845, XP002032093
K.L. PURANAM AND P.J. BLACKSHEAR:
"Cloning and characterization of RECQL, a potential human homologue of the Escherichia coli DNA helicase RecQ"
see the whole document
--- | 1-21 |
| P,X | SCIENCE (WASHINGTON D C) 272 (5259). 1996. 258-262. ISSN: 0036-8075, XP002032094
YU C-E ET AL: "Positional cloning of the Werner's syndrome gene."
see the whole document
--- | 1,2,10 |
| P,X | TRENDS IN GENETICS 12 (8). 1996. 283-286. ISSN: 0168-9525,
August 1996, XP002032095
LOMBARD D B ET AL: "Cloning the gene for Werner syndrome: A disease with many symptoms of premature aging."
see the whole document
--- | 1,2,10,
15,16 |
| P,X | HUMAN MOLECULAR GENETICS 5 (12). 1996. 1111-1114. ISSN: 0168-9525,
December 1996, XP002032096
W. K. CHEN ET AL: "Cloning and characterization of the Werner syndrome gene."
see the whole document
--- | 1,2,10 |

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 96/20785

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|---|-----------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| T | AMERICAN JOURNAL OF HUMAN GENETICS 60 (2).
1997. 330-341. ISSN: 0002-9297,
February 1997, XP000674411
YU C-E ET AL: "Mutations in the consensus
helicase domains of the Werner syndrome
gene."
see the whole document
--- | 1-21 |
| T | GENOMICS 41 (2). 1997. 298-300. ISSN:
0888-7543, XP000674356
IMAMURA O ET AL: "Cloning of a mouse
homologue of the human Werner syndrome
gene and assignment to 8A4 by fluorescence
in situ hybridization."
see the whole document
----- | 1-21 |

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/05330

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : A61K 48/00; C12N 5/00, 5/06, 5/08, 5/10, 5/12, 5/16, 5/22

US CL : 435/240.1, 240.2, 172.3, 320.1; 424/93.21; 935/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/240.1, 240.2, 172.3, 320.1; 424/93.21; 935/62

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAS ONLINE, APS, bystander#, migration, msc, mc, gene#, dna#, cdna#, rna#, mrna#, gene, thera?

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|--|-----------------------|
| Y | WO 93/02556 [FREEMAN ET AL] 18 Februray 1993, see entire document. | 1-33 |
| Y | Science, Volume 256, issued 12 June 1992, K.W. Culver, "In vivo gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors", pages 1550-1552, see entire document. | 1-33 |
| Y | Cancer Research, Volume 53, issued 01 January 1993, Z. Ram et al, "In situ retroviral-mediated gene transfer for the treatment of brain tumors in rats", pages 83-88, see entire document. | 1-33 |

BEST AVAILABLE COPY

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

| | |
|---|---|
| * Special categories of cited documents | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| *A* document defining the general state of the art which is not considered to be of particular relevance | *V* document of the state of the art which is not considered to be of particular relevance |
| *E* earlier document published on or after the international filing date but prior to the priority date claimed | *X* document of the state of the art which is not considered to be of particular relevance |
| *P* document published prior to the international filing date but later than the priority date claimed | *G* document member of the same patent family |

Date of the actual completion of the international search

24 JULY 1995

Date of mailing of the international search report

03 AUG 1995

Name and mailing address of the ISA-US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

JAMES MARTINELL

Telephone No. (703) 308-0196

International application No
PCT/US95/05330

International application No
PCT/US95/05330

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|--|-----------------------|
| Y | Proceedings of the National Academy of Sciences, USA, Volume 85, issued October 1988, E. Borrelli et al, "Targeting of an inducible toxic phenotype in animal cells", pages 7572-7576, see entire document. | 1-33 |
| Y | Journal of Experimental Medicine, Volume 172, issued October 1990, B. Gansbacher et al, "Interleukin 2 gene transfer into tumor cells abrogates tumorigenicity and induces protective immunity", pages 1217-1224, see entire document. | 1-33 |
| Y | Science, Volume 254, issued 01 November 1991, P.T. Golumbek et al, "Treatment of established renal cancer by tumor cells engineered to secrete interleukin-4", pages 713-716, see entire document. | 1-33 |
| Y | Journal of Experimental Medicine, Volume 177, issued May 1993, L.E. Minasi et al, "The selective ablation of interleukin 2-producing cells isolated from transgenic mice", pages 1451-1459, see entire document. | 1-33 |

BEST AVAILABLE COPY

BEST AVAILABLE COPY